VOLUME 37 MAY, 1951 NUMBER 5

# PROCEEDINGS

OF THE

# NATIONAL ACADEMY OF SCIENCES

OF THE

UNITED STATES OF AMERICA

EDITORIAL BOARD

LINUS PAULING, *Chairman* EDWIN B. WILSON, *Managing Editor*
FRED. E. WRIGHT, *Home Secretary* ROGER ADAMS, *Foreign Secretary*
DOUGLAS WHITAKER, *Chairman of the National Research Council*

J. W. BRAMS
R. E. CLELAND
W. W. RUBEY
HARLOW SHAPLEY
F. E. TERMAN
G. W. BEADLE
E. A. DOISY
G. C. EVANS
C. H. GRAHAM
J. T. PATTERSON
GREGORY BREIT
L. C. DUNN
C. S. MARVEL
J. VON NEUMANN
W. H. TALIAFERRO

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office: Harvard School of Public Health, Boston 15
Home Office of the Academy: Washington 25, D. C.

Entered as Second-Class matter, June 21, 1920, at the post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in section 1103, act of October 3, 1917. Authorized June 12, 1920.

Published Monthly Price per Annum, $7.50

## OFFICERS OF THE ACADEMY

PRESIDENT ..... D. W. BROWN
VICE PRESIDENT ..... EDWIN B. WILSON
FOREIGN SECRETARY ..... ROGER ADAMS
HOME SECRETARY ..... FRED. E. WRIGHT
TREASURER ..... W. J. ROBBINS

## INFORMATION TO SUBSCRIBERS

Checks covering subscriptions to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES at the rate of $7.50 per annum should be made payable to the National Academy of Sciences, and sent to the Treasurer, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C. Single numbers $0.75. These prices effective January, 1961.

---

Microfilm copies of the PROCEEDINGS are available to subscribers to the paper edition. The microfilm is a positive, furnished on metal reels, suitably labeled and is distributed at the end of the volume year. Inquiries concerning purchase should be directed to University Microfilms, 313 N. First St., Ann Arbor, Michigan.

---

Past publications of the NATIONAL ACADEMY OF SCIENCES from 1863 to 1926 and of the NATIONAL RESEARCH COUNCIL from 1916 to 1925 are listed in the Index to the First Ten Volumes of these PROCEEDINGS. This Index appeared as Part 2 of Volume 13, No. 1, January, 1927, and is for sale separately at the price of $1.50. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

PAST volumes of the PROCEEDINGS may be obtained at five dollars per volume unless the sale of the volume would break a complete set of Volumes I to XXX. Single numbers may be obtained for fifty cents except where the sale of such numbers would break up a complete volume. Only sixty complete sets of the PROCEEDINGS are available for sale—Volumes I to XXX price $200.00. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

---

The following publications are issued by the NATIONAL RESEARCH COUNCIL. Orders and inquiries should be addressed: Publication Office, National Research Council, 2101 Constitution Ave., Washington 25, D. C.

(A) THE BULLETIN OF THE NATIONAL RESEARCH COUNCIL published at irregular intervals. Individual numbers priced variously. A discount of 20% from the list price will be given to all who subscribe for the series as issued, direct from the Publication Office.

(B) THE REPRINT AND CIRCULAR SERIES OF THE NATIONAL RESEARCH COUNCIL, individual numbers variously priced. A discount of 20% from the list price will be given to all who subscribe for the series as issued direct from the Publication Office.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 37 May 15, 1951 Number 5

# *ATOMIC COORDINATES AND STRUCTURE FACTORS FOR TWO HELICAL CONFIGURATIONS OF POLYPEPTIDE CHAINS*

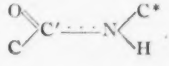
By Linus Pauling and Robert B. Corey

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology, Pasadena, California

Communicated March 31, 1951

During recent years we have been gathering information about interatomic distances, bond angles, and other properties of simple substances related to proteins, and have been attempting to formulate configurations of the polypeptide chain that are compatible with this information and that might constitute a structural feature of proteins. We have reported the discovery of two helical configurations that satisfy these conditions.[1,2] In the following paragraphs we discuss the atomic positions for these configurations, and their form factors for diffraction of x-rays in the equatorial direction.

*The γ Helix.*—Let us first discuss the 5.1-residue helix. This configuration is obtained by coiling a polypeptide chain into a helical form, in such a way that the planar amide groups,




are in the trans configuration (the carbonyl group being almost directly opposed to the imino group), and each amide group forms hydrogen bonds with the fifth more distant group in each direction along the chain. The structure is represented diagrammatically in figure 1, and a drawing of it has been recently published.[2] We base our discussion on the values of interatomic distances and bond angles given in figure 4; these differ from those described earlier only in the change from 120° to 123° for the angle C′—N—C*.

The fifth amide group beyond a given group in the helix is nearly directly above it, and if the hydrogen bonds determine the orientation of the plane of the amide groups there seems to be no reason for this plane not to be parallel to the axis of the helix. The following calculations are made

with this assumption. It is found that the rise per residue—the magnitude of the translation associated with the rotatory translation that converts one residue into the adjacent one along the chain—is determined nearly exactly by the length of the hydrogen bond. For N—H···O distance 2.72 A the rise per residue is 0.97 A, for 2.78 A it is 0.98 A, and it continues to increase linearly by about 0.002 A for each 0.010 A increase in length of the hydrogen-bond distance. The number of residues per turn, on the other hand, is determined essentially by the N—C—C′ angle of the $\alpha$ carbon

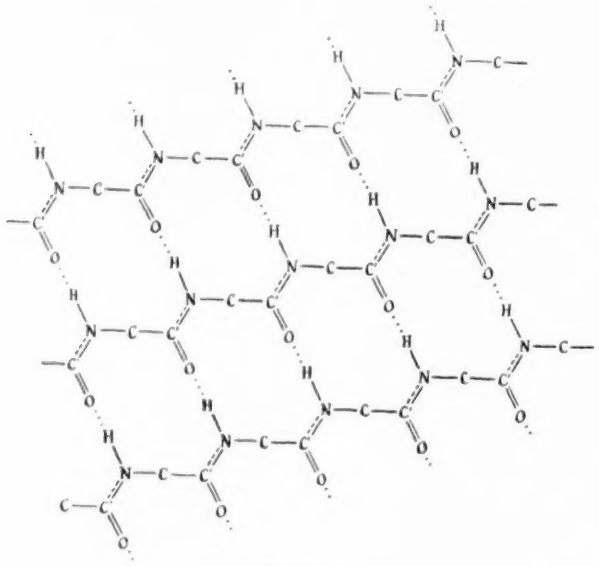


FIGURE 1

Diagrammatic representation of the 5.1-residue helical configuration of the polypeptide chain.

atom. For 21 residues in 4 turns (5.25 residues per turn) this angle has the value 111.2°, for 26 residues in 5 turns (5.20) its value is 110.6°, for 31 residues in 6 turns (5.17) 110.1°, and for 36 residues in 7 turns (5.14 residues per turn) 109.8°. These values are all within a reasonable range for this angle; the existing experimental evidence, for scores of molecules containing a tetrahedral carbon atom, suggests 110° as the best value for the angle, but a change by as much as 1° would introduce such a small amount of strain energy, in any case, that it should be allowed. There is, of course, no reason in an isolated molecule of a polypeptide or protein for the number of residues per turn to be rational. In a crystal, however, the

intermolecular forces might constrain the molecule to assume the symmetry of a position in the crystal. The most likely eventuality is that the cylindrical molecules would take up a hexagonal close-packed arrangement, and that the molecules would assume a sixfold screw axis. This is provided from among the foregoing possibilities only by the configuration with 36 residues in 7 turns, which has a 36-fold screw axis, with a sixfold screw axis contained within its symmetry group. We present in table 1 calculated atomic coordinates for this case, for which the angle of rotation of the fundamental rotatory-translational operation is exactly 70°. The coordinates have been calculated for a rise per residue 0.98 A, corresponding to the

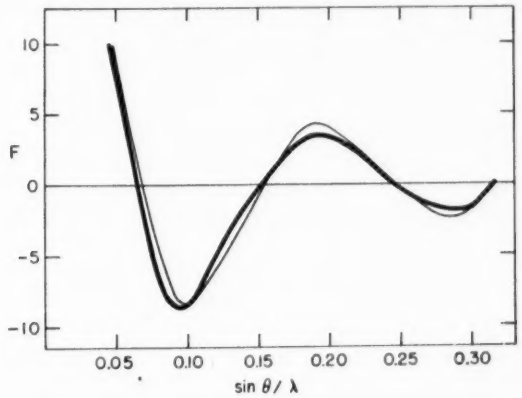



FIGURE 2

X-ray form factor for the 5.1-residue helix, calculated for equatorial reflections, and for cylindrical symmetry. The light line represents the scattering due to the four main-chain atoms of the amide group, and the heavy line includes the scattering of one $\beta$ carbon atom per residue.

distance N—H···O equal to 2.78 A. The rise per turn is then 5.04 A, and the identity distance along the helical axis is 35.28 A.

The form factor for x-ray scattering for this molecule could, of course, be calculated from the coordinates of the atoms. A very good approximation for the equatorial reflections can be made by assuming cylindrical symmetry about the axis. The structure factor is then $F = \sum_i f_i J_0 (4\pi\rho_i \sin\theta/\lambda)$, in which $J_0$ is the Bessel function of order zero with the indicated argument, and $\rho_i$ is the radius of the $i$th atom, in cylindrical coordinates, as given in table 1. (In this table there are given both cartesian coordinates, $x$, $y$, $z$, and cylindrical coordinates, $\rho$, $\phi$, $z$, the latter being relative to the axis of the helix.) The form factor $F$ as calculated by this method with use of the atomic form factors given in the International Tables for the

Determination of Crystal Structures is shown in figure 2, both for the residue alone, corresponding to polyglycine, and for a molecule with a $\beta$ carbon atom in each residue. The two form factors do not differ very

TABLE 1

ATOMIC COORDINATES FOR THE 36-RESIDUE 7-TURN $\gamma$ HELIX

$x$, $y$, $z$, $\rho$ IN A

| ATOM | $x$ | $y$ | $z$ | $\rho$ | $\phi$ |
|---|---|---|---|---|---|
| C | 0.00 | 0.00 | 0.00 | 3.22 | 0.0° |
| C′ | 1.52 | 0.00 | −0.15 | 2.66 | 27.8° |
| O | 2.05 | 0.00 | −1.26 | 2.65 | 39.3° |
| N | 2.23 | 0.00 | 0.97 | 2.67 | 43.2° |
| C* | 3.70 | 0.00 | 0.98 | 3.22 | 70.0° |
| $\beta$C | −0.50 | −0.74 | 1.26 | 4.11 | 0.7° |
| or | | | | | |
| $\beta$C | −0.52 | −0.72 | −1.26 | 4.11 | −0.7° |
| Axis | 1.85 | 2.64 | ... | 0.00 | ... |

much. The square of $F$, which determines the intensity of reflection of x-rays, is shown in figure 3. It is seen that we predict that the equatorial

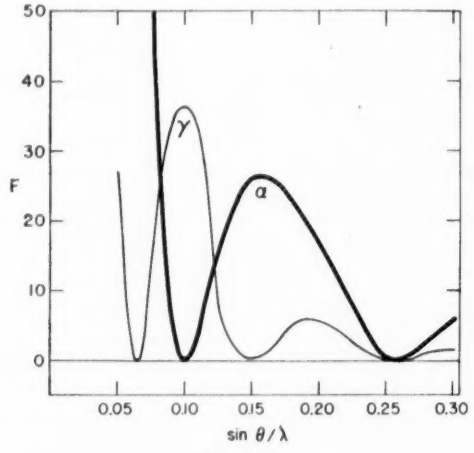



FIGURE 3

The square of the form factor for the 5.1-residue helix ($\gamma$) and the 3.7-residue helix ($\alpha$), for equatorial reflections.

reflections from an array of molecules with this structure would be weak in the regions of interplanar distance 7.7 A, 3.3 A, and 2.0 A, and strong at

about 5.0 A and 2.4 A. A discussion of the question of possible existence of protein molecules with this structure will be given in a later paper.

*The α Helix.*—The α helix is a configuration of the polypeptide chain in which each amide group forms hydrogen bonds with the amide groups removed by three from it in either direction along the chain. The α helix and the γ helix are the only helical configurations in which the residues are all equivalent and in which the stereochemical requirements, including formation of intramolecular hydrogen bonds, are satisfied. The structure is represented diagrammatically in figure 4. With the structural parameters described above, we find that a translation per residue of 1.47 A corresponds to a hydrogen-bond distance 2.75 A, and that the translation increases by 0.01 A for every 0.03 A increase in the hydrogen-bond distance. The number of residues per turn is fixed primarily by the bond angle at the α carbon atom; it varies from 3.60 for bond angle 108.9° to 3.67 for bond angle 110.8°. α-Helixes with these numbers of residues per turn have been found (see the following paper[3]) to explain the x-ray reflections[4] from highly oriented fibers of poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate, respectively. The first ratio corresponds to 18 residues in 5 turns and the second to 11 residues in 3 turns.

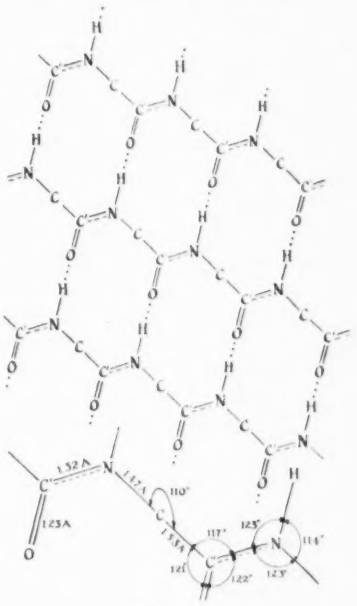



FIGURE 4

Diagrammatic representation of the 3.7-residue helical configuration of the polypeptide chain.

We have chosen to present parameters for the 18-residue 5-turn helix, which has a sixfold screw axis that would be stabilized in hexagonal packing. The parameters given in table 2 correspond to this helix, with a translation of 1.50 A per residue (5.40 A per turn, 27.0 A identity distance along the axis), these being the dimensions found for the polymethylglutamate. The corresponding hydrogen-bond distance is 2.86 A.

The structure factor for equatorial reflections calculated for cylindrical symmetry is shown in figure 5, and the corresponding intensity curve in figure 3. It is seen that the equatorial reflections should be very weak at about 5.0 A and 2.0 A, and strong at about 3.4 A. The correlation be-

tween theory and experiment for synthetic polypeptides and for proteins with the α-keratin structure is discussed in following papers of this series.

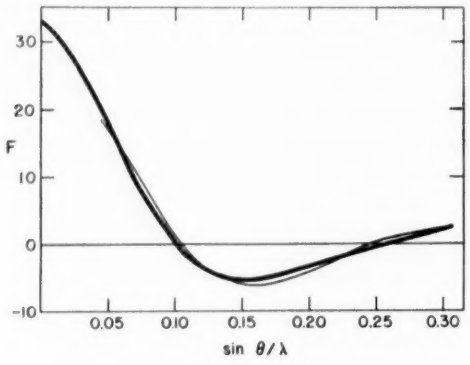



FIGURE 5

X-ray form factor for the 3.7-residue helix, calculated for equatorial reflections, and for cylindrical symmetry. The light line represents the scattering due to the main-chain atoms only, and the heavy line includes also the scattering of one β carbon per residue.

TABLE 2

ATOMIC COORDINATES FOR THE 18-RESIDUE 5-TURN α HELIX

$x$, $y$, $z$, $\rho$ IN A

| ATOM | $x$ | $y$ | $z$ | $\rho$ | $\phi$ |
|---|---|---|---|---|---|
| C | 0.00 | 0.00 | 0.00 | 2.29 | 0.0° |
| N | 1.16 | 0.00 | 0.89 | 1.59 | 27.8° |
| C′ | 2.42 | 0.00 | 0.44 | 1.61 | 73.8° |
| O | 2.69 | 0.00 | −0.76 | 1.74 | 82.0° |
| C* | 3.52 | 0.00 | 1.50 | 2.29 | 100.0° |
| βC | −1.33 | 0.20 | 0.76 | 3.34 | −17.6° |
| or | | | | | |
| βC | −0.03 | −1.34 | −0.76 | 3.34 | 17.6° |
| Axis | 1.76 | 1.47 | ... | 0.00 | ... |

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The United States Public Health Service.

* Contribution No. 1550.

[1] Pauling, L., and Corey, R. B., *J. Am. Chem. Soc.*, **72,** 5349 (1950).

[2] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37,** 205 (1951).

[3] Pauling, L., and Corey, R. B., *Ibid.*, **37,** 241 (1951).

[4] Bamford, C. H., Hanby, W. E., and Happey, F., *Proc. Roy. Soc.*, **A205,** 30 (1951).

# THE STRUCTURE OF SYNTHETIC POLYPEPTIDES

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

In a preliminary communication last year[1] we stated that there are only two helical configurations of polypeptide chains in which the residues are all equivalent and intramolecular hydrogen bonds are formed, and in which the interatomic distances, bond angles, and other structural features, especially the coplanarity of the conjugated amide system, are as required by earlier work in these Laboratories on amino acids, simple peptides, and other substances related to proteins. These two helical configurations were described in detail in a later paper[2] and it was mentioned that there is evidence that they occur in $\alpha$ keratin, $\alpha$ myosin, supercontracted keratin and myosin, and other fibrous proteins, and also constitute an important structural feature of hemoglobin and other globular proteins.[3] In the following paragraphs we discuss evidence that one of the helical structures is assumed also by synthetic polypeptides.

Oriented films and fibers of several synthetic polypeptides have been prepared and examined by x-ray diffraction by Bamford, Hanby, and Happey,[4] and the oriented films have been investigated with polarized infrared spectroscopy by Ambrose and Elliott.[5] The best photographs were given by poly-$\gamma$-methyl-L-glutamate and poly-$\gamma$-benzyl-L-glutamate. The x-ray data indicate an identity distance in the direction of the fibers of 5.50 A for the first substance, and a slightly larger value for the second substance. The dichroism observed for the N—H stretching and bending infrared absorption bands and for the C=O stretching band indicates that these groups are oriented nearly parallel to the fiber axis, and the conclusion is accordingly drawn that intramolecular hydrogen bonds are formed nearly parallel to this axis. These authors have interpreted all of their data as providing strong support for the $\alpha_{II}$ structure shown in figure 1. This structure was first proposed by Huggins[6] and has been discussed by Zahn,[7] Simanouti and Mizushima,[8] and Ambrose and Hanby.[9]

The $\alpha_{II}$ structure must, however, be rejected. In the structure as discussed by Bamford and coworkers the amide group is not assigned a planar configuration. There is, in fact, very strong theoretical and experimental evidence that the carbon-nitrogen bond has a large amount of double-bond character, and that the four atoms adjacent to these two atoms must form a coplanar system with them. The carbon-nitrogen distance is 1.32 A, which is 0.15 A less than the single-bond distance between these atoms. This shortening corresponds to about 50 per cent double-bond character,

which is great enough to lead to effective coplanarity of the system. (Ambrose and Elliott[5] say "It remains to be shown whether resonance in polypeptides and proteins effectively limits the rotation about the C—N bond, and we regard the matter as by no means settled." In fact, however, our present knowledge of structural chemistry is such that there can be no doubt on this point: non-planar configurations of this group must surely be accompanied by pronounced instability.) In our investigation of helical configurations of the polypeptide chain[2] we found it impossible to construct an acceptable configuration resembling $\alpha_{II}$. The closest approximation to an acceptable configuration that can be constructed, in which the hydrogen atom of the N—H group is brought to within 1.8 A of the carbonyl oxygen atom, is unsatisfactory because the N—H···O angle differs from a straight angle by about 70°, and thus this configuration would be presumed

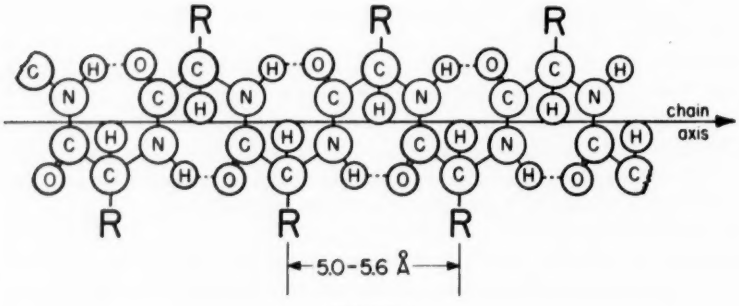



FIGURE 1

The $\alpha_{II}$ structure for the polypeptide chain, discussed by Bamford, Hanby, and Happey, and by Ambrose and Elliott.

not to correspond to a reasonably strong hydrogen bond. It was for this reason that the $\alpha_{II}$ structure was eliminated in our earlier considerations. There are also several other arguments against the $\alpha_{II}$ structure. Ambrose and Elliott mention that their model gives a dichroic ratio of 1.41:1 for the C=O stretching calculated for perfectly oriented molecules, and that the observed dichroism in poly-$\gamma$-benzyl-L-glutamate, 2.6:1, is considerably greater than this value. Moreover, the $\alpha_{II}$ structure corresponds to an extension of only approximately 40 per cent during the $\alpha \rightarrow \beta$ transformation. Ambrose and Elliott suggest that the observed reversible extension of wool, 100 per cent or more, is not to be interpreted as necessarily requiring a similar reversible extension from the $\alpha$ to the $\beta$ configuration of the polypeptide chain, because there is present some amorphous material in wool; but it seems to us, as also to Astbury,[10] very unlikely that such a great discrepancy could exist. In addition, it may be pointed out that the

synthetic polypeptides seem to be hexagonal or closely pseudohexagonal in structure, and the flat $\alpha_{II}$ configuration provides no explanation of this fact.

All of these difficulties are overcome if it is assumed that the synthetic polypeptides have the configuration of the third-amide hydrogen-bonded helix which we have described. In this helix the distance per residue along the helical axis is predicted to be approximately 1.50 A and there are approximately 3.7 residues per turn. The translation along the axis per turn is thus predicted to be about 5.5 A. This is in excellent agreement with the fiber axis translations reported by Bamford, Hanby, and Happey,

TABLE 1

X-RAY DATA FOR POLY-γ-METHYL-L-GLUTAMATE
HEXAGONAL UNIT WITH $a_0$ = 11.96 A, $c_0$ = 27.5 A

Equator:

| $HI \cdot L$ | $d_{obs.}$ | $d_{calc.}$ | $F_1$[a] | $F_2$ | $I_{calc.}$ | $I_{obs.}$ |
|---|---|---|---|---|---|---|
| 10·0 | 10.35 A | 10.35 A | 18.5 | 16.2 | 360 | VVS |
| 11·0 | 5.98 | 5.98 | 5.1 | 6.2 | 29 | S |
| 20·0 | 5.22 | 5.18 | 1.3 | 3.2 | 7 | M |
| 21·0 | 3.89 | 3.91 | −4.0 | −4.8 | 22 | S |
| 30·0 | 3.45 | 3.45 | −5.0 | −7.1 | 21 | S |
| 22·0 | 3.00 | 2.99 | −5.0 | −6.1 | 14 | S |
| 31·0 | 2.87 | 2.87 | −4.9 | −1.8 | 2 | W |

Second hyperbola

| $HI \cdot L$ | $d_{obs.}$ | $d_{calc.}$ |
|---|---|---|
| 11·2 | 5.51 A | 5.48 A |
| 20·2 | 4.88 | 4.87 |

Fifth hyperbola

| $HI \cdot L$ | $d_{obs.}$ | $d_{calc.}$ |
|---|---|---|
| 10·5 | 4.82 A | 4.86 A |
| 11·5 | 4.05 | 4.05 |
| 20·5 | 3.75 | 3.77 |

Sixth hyperbola: a polar arc with $d_{obs.}$ = 4.43 A; $d_{calc.}$ for $\{10.6\}$, 4.20 A.

[a] Calculated for cylindrically symmetrical distributions of atomic centers, with radii from helical axis 2.29 A for C, 1.59 A for N, 1.61 A for C′, 1.74 A for O, and 3.34 A for βC.

5.50 A for poly-γ-methyl-L-glutamate and slightly larger for the benzyl ester. Moreover, the helical molecules would be expected to pack together essentially as would cylinders, in a hexagonal packing, and the predicted interplanar distances for the hexagonal lattice agree well with those observed, and shown in tables 1 and 2.

Bamford, Hanby, and Happey interpreted the data for the benzyl ester in terms of an orthorhombic unit with axes 10.35 A, 5.98 A, and 5.50 A. The number of molecules in this unit calculated from the density 1.34 is 1.9. The authors note that the first two axial lengths are in the ratio $\sqrt{3}:1$, which is compatible with hexagonal symmetry, and they ask to what extent the x-ray results are consistent with a helical configuration of the mole-

cules. They point out that a hexagonal unit with $a_0 = 11.96$ A and $c_0 = 5.50$ A and containing three amino-acid residues, corresponding to a three-fold helix, would have a density 1.06 g cm$^{-3}$, which is too small. Our helix, with about 3.7 residues per turn, would lead, however, to the density 1.29, which is acceptable. A similarly satisfactory value of the density, 1.3, is also calculated for the corresponding hexagonal unit for the benzyl ester, with interplanar distances given in table 2.

Although most of the reflections observed by Bamford, Hanby, and Happey are accounted for by the orthorhombic unit or the hexagonal unit

TABLE 2

X-RAY DATA FOR POLY-γ-BENZYL-L-GLUTAMATE

PSEUDO-ORTHORHOMBIC UNIT WITH $a_0 = 25.0$ A, $b_0 = 17.3$ A, $c_0 = 14.42$ A

Equator:

| *hkl* | $d_{calc.}$ | $d_{obs.}$ | $F_1$ | $F_2$ | $I^2_{calc.}$ | $I_{obs.}$ |
|---|---|---|---|---|---|---|
| 100 | 25.0 | ... | 0 | 1.3 | 0.5 | W |
| 001 | 14.4 | ... | 0 | 1.1 | 0.4 | W |
| 101, 200 | 12.5 | 12.6 | 21.5 | 17.8 | 165 | VVS |
| 002, 301 | 7.21 | 7.21 | 9.5 | 11.1 | 35 | VS |
| 202, 400 | 6.25 | 6.21 | 6.0 | 9.0 | 20 | S |
| 103, 402, 501 | 4.72 | 4.87 | −0.8 | −2.2 | 2 | M (diffuse) |
| 303, 600 | 4.17 | 4.07 | −3.8 | −7.4 | 9 | S (diffuse) |
| 004, 602 | 3.61 | 3.85 | −4.4 | −4.8 | 4 | |
| 204, 503, 701 | 3.49 | ... | −4.8 | +0.8 | 0.2 | ... |

First hyperbola: faint reflections at 16 A.
Second hyperbola: faint reflections at 8 A.
Third hyperbola:

| *hkl* | $d_{calc}$ | $d_{obs.}$ | $I_{obs.}$ |
|---|---|---|---|
| 151, 250 | 5.24 A | 5.27 A | VS |
| 052, 351 | 4.50 | 4.48 | S (diffuse) |
| 252, 450 | 4.24 | 4.08 | |
| 153, 452, 551 | 3.65 | 3.60 | |

with $c_0 = 5.50$ A, some reflections which they could not explain were reported by them. They concluded that these reflections, observed for both of the esters, are due to a second phase in each specimen, because the reflections seemed not to have any obvious relation to those that had been indexed. The second phase would have to be assumed to have parallel orientation to the main phase. It has seemed to us likely that these weak reflections indicate the presence of a larger unit for each crystal. A larger unit of structure would, of course, be predicted on the basis of our 3.7-residue helix, inasmuch as a true identity distance along the helical axis could occur only after a number of turns. Even though the interatomic interactions that are operative in stabilizing the helix would not be expected,

in a non-crystalline phase, to lead to a rational number of residues per turn, the forces operating during crystallization might well be great enough to produce a small torque in the helix, such as to cause it to assume a configuration with a rational number of residues per turn. The three simplest rational helixes of the 3.7-residue hydrogen-bonded type are those with 11 residues in 3 turns (3.67 residues per turn), 15 residues in 4 turns (3.75 residues per turn), and 18 residues in 5 turns (3.60 residues per turn). We have found that the weak reflections for the methyl ester can be accounted for on the basis of the 18-residue 5-turn helix, and those for the benzyl ester on the basis of the 11-residue 3-turn helix. Plans of these helixes are shown as figures 2 and 3.

The 18-residue 5-turn helix has a sixfold screw axis of symmetry, and it would accordingly be expected that these helical molecules would arrange themselves side by side in hexagonal packing in such a way that the unit

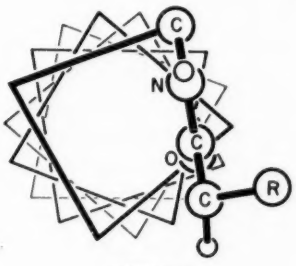



FIGURE 2

Plan of the 18-residue 5-turn helix.

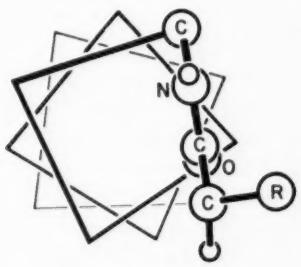



FIGURE 3

Plan of the 11-residue 3-turn helix.

of structure would contain only one 18-residue segment of one helix. The predicted dimensions of the hexagonal unit of structure would, then, be $a_0 = 11.96$ A, $c_0 = 27.5$ A. All of the data reported by Bamford, Hanby, and Happey are accounted for by this unit, as shown in table 1. The two reflections in the second hyperbola and the one in the sixth hyperbola are those which are not accounted for by the smaller unit. The earlier authors described the two reflections in the second hyperbola as corresponding to a $c$-axial length of about 13.3 A, which is close to half of our value of $c_0$. The reflection at 4.43 A is a polar arc, which probably could not be measured very accurately.

The determination of the structure of the crystals of the methyl ester would involve only the determination of the orientation of one of the helixes and of the positions of the atoms in the side chains. Simple considerations show that the observed intensities are approximately those predicted for a structure of this sort.

We have calculated intensities of the equatorial reflections for this hexagonal structure, with the four peptide atoms CNCO and the $\beta$ carbon atom distributed over cylindrical surfaces, the structure factor then being $F_{HI\cdot L} = \sum_i f_i J_0(2\pi\rho_i/d_{HI\cdot L})$; here $f_i$ is the atomic structure factor for the $i$th atom, $\rho_i$ is the distance of the atom from the helical axis, as given by our calculations,[2] and $J_0$ is the Bessel function of order zero. The values found are included in table 1, as $F_1$. It is seen that there is rough general agreement. Our calculations have included only five of the ten heavy atoms per residue; and although it may be predicted that the contribution of the

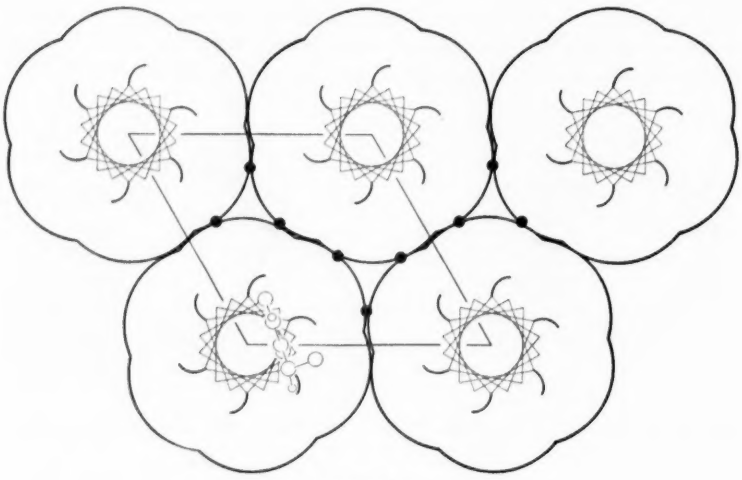


FIGURE 4

Proposed structure of poly-$\gamma$-methyl-L-glutamate.

other five (side-chain) atoms would be less, because of greater mutual interference, than that of the five main-chain atoms, it should be significant, and can explain the observed medium intensity of $\{20.0\}$ and weak intensity of $\{31.0\}$. That this can be done is seen by comparison with $F_2$, and with $I_2$, the corresponding intensities $pLPF_2^2$, in which $p$ is the frequency factor, $L$ the Lorentz factor, and $P$ the polarization factor. In this calculation it is assumed that side-chain atoms are grouped about the 3-fold screw axes, in the positions $xyz$, with $xy = {}^3/_7\,{}^6/_7, {}^3/_7\,{}^4/_7, {}^1/_7\,{}^4/_7, {}^4/_7\,{}^1/_7, {}^6/_7\,{}^3/_7$, and ${}^4/_7\,{}^3/_7$. These positions are indicated by solid circles in figure 4, which represents the proposed structure for the methyl ester polymer. One atom per residue has been placed in these positions; the calculated intensities correspond

to one oxygen atom and two carbon atoms for each asymmetric unit of three residues.

It is interesting to note that a reasonable explanation of the observed intensities of the equatorial reflections can be obtained with neglect of 40 per cent of the heavy atoms in the crystal. We believe that this phenomenon, which seems to be rather general for proteins and related substances, is due to the larger mutual interference of the rays scattered by the side-chain atoms than of those scattered by the main-chain atoms. The side-chain atoms, of which there are 90 in the unit, are located in 15 sets of 6-fold positions, presumably at 15 different values of the radius from the helical axis; whereas the 90 main-chain atoms (including the $\beta$ carbon atoms) are more regularly arranged, corresponding to the pseudo 18-fold screw axis of the helix, and only five values of the radius are represented.

The anomalous reflections reported for the benzyl ester are described as lying on diffuse hyperbolas at 16 A and 8 A. We interpret these reflections as resulting from the presence of the 11-residue 3-turn helix, for which the value 16.8 A for $c_0$ would be predicted, on the assumption that the residue distance is the same as in the methyl ester, 1.53 A. This corresponds to 5.60 A per turn; Bamford, Hanby, and Happey suggested 5.76 A or a somewhat smaller value for the translation along the fiber axis, and we have found the data to correspond to $3 \times 5.76 = 17.3$ A, or 1.57 A per residue. The increase of 0.04 A over the methyl ester may well be due to van der Waals repulsion of the side chains.

The 11-residue 3-turn helix does not have a 6-fold axis or 3-fold axis, and accordingly it would not be expected to form hexagonal crystals with one helix per unit. The larger residue weight of the benzyl ester (219 in place of 143 for the methyl ester) would lead to a larger pseudo-hexagonal unit, with $a_0$ about 14.4 A, rather than 10.96 A (table 2). (The values given in table 2 are those obtained with the x-ray beam through the edge of an oriented film; closely similar values were also obtained with the x-ray beam perpendicular to the film, and with oriented fibers.)

Let us consider ways in which an asymmetric helix might be surrounded by others. We assume that the helixes are all equivalent, and that they are in contact with one another in a simple way. The six vectors from each asymmetric helix to its six neighbors are of six different kinds, representing six kinds of interaction. These might consist of three pairs, with the two of each pair differing only in polarity. It is found that there are only three arrangements of the three pairs in which the helixes are equivalent, and that none of them explains the occurrence of the 25-A reflection and the 14.4-A reflection for the benzyl polymer. It is possible, however, that two adjacent helixes interact in a way without polarity, to form a doublet. This would occur, for example, if one helix were related to the other by a 2-fold screw axis. There are many ways of arranging these doublets in

a crystal; one of the simplest, shown in figure 5, has monoclinic (pseudo-hexagonal) symmetry, and for the benzyl ester corresponds to axes $a_0 = 25.0$ A, $b_0 = 17.3$ A, $c_0 = 14.42$ A, and $\beta = 90°$. This unit accounts for all of the reflections observed for this polymer.

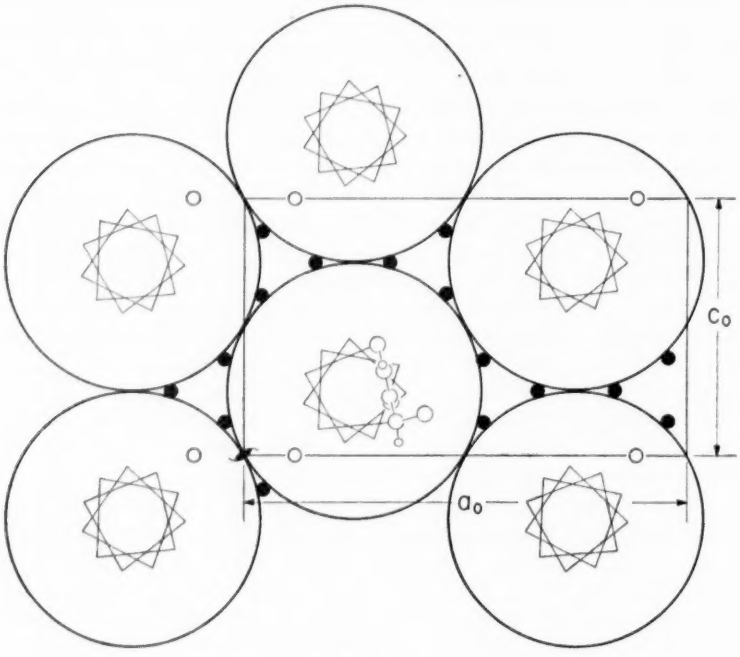



FIGURE 5

Proposed structure of poly-$\gamma$-benzyl-L-glutamate.

The intensities of all equatorial reflections can be roughly accounted for by consideration of only the five main-chain atoms, cylindrically distributed, as described above, about helical axes with coordinates $x$, $z$, and $\bar{x}$, $\bar{z}$, with $x = {}^1/_4$ and $z = {}^1/_4$, and one additional atom per residue, in the same positions as assumed for the methyl ester. The weak reflections $\{100\}$ and $\{001\}$ can be accounted for by any one of various slight distortions from the ideal hexagonal structure. The values of $F_1$ in table 2 are the calculated form factors for the five main-chain atoms per residue, and those of $F_2$ include the contributions of additional atoms in the positions assumed for the methyl ester (shown by full circles in figure 5). The scattering power in this position has been taken as 50 per cent greater than for

the methyl ester. An additional concentration of electrons equivalent to one oxygen atom per unit of eleven residues has also been placed at each of the positions $x$, $z = \pm 1/8$, 0 (shown by open circles in the figure). This small additional concentration of electrons, representing the deviation from the hexagonal symmetry of the methyl ester, accounts for the faint reflections $\{100\}$ and $\{001\}$. The close approximation of the structure to hexagonal packing of circular cylinders is made evident not only by the close approximation of the axial ratio $a/c$ to $\sqrt{3}$, but also by the fact that the intensities of the two reflections requiring the larger unit can be explained by the scattering power of two heavy atoms, out of the 330 in the unit cell. Any one of many alternative small deviations of the distribution of scattering power from the ideal distribution with hexagonal symmetry would, of course, account for these observed weak reflections.

It seems likely that a poly-L-glutamate helix would be more stable with one screw sense (right-handed or left-handed) than with the other, and that helixes of only one kind are formed in significant number in the process of folding. A hexagonal array of helixes might then consist of equal numbers with positive and with negative orientation and essentially random distribution over the lattice points, equal numbers in an ordered array, only helixes with one orientation, or (much less likely) an ordered array in a ratio other than 1:1. For poly-$\gamma$-methyl-L-glutamate the detail of the photographs and the presence of only one helix per unit cell strongly suggest that the crystals contain helixes with only one orientation, and that segregation has occurred in crystallization. For the poly-benzyl ester the two-molecule unit also suggests that the helixes all have the same orientation; the strong reflections all have $h + k + l$ even, as required for an approximately body-centered unit. The possibility that segregation of positively and negatively oriented molecules occurs during crystallization suggests that annealing the specimens might improve the photographs.

Bamford, Hanby, and Happey also investigated several copolymers, and found from x-ray investigation that some tended to crystallize with the $\alpha$ configuration and some, especially those containing a large fraction of glycine residues, with the $\beta$ configuration, involving sheets of extended polypeptide chains with lateral hydrogen bonds. They found that the $\alpha$ phase was oriented much more readily by stretching than the $\beta$ phase, and attributed this difference to the presence of intramolecular hydrogen bonds in the $\alpha$ phase molecules. This feature remains unchanged by our attribution of our helical structure to the $\alpha$ phase. We explain the stability of the $\beta$ phase for the glycine polymers by the stability of the lateral hydrogen bonds that can be formed; we have noted that with other residues there is serious steric interference between side chains, for the $\beta$ configuration, leading to increase in the hydrogen-bond distance by 0.2 A, and a corresponding decrease in stability, relative to the $\alpha$ helix.

Bamford, Hanby, and Happey report interplanar distances leading to fiber-axis pseudo identity distances of 5.75 A for each of the three copolymers poly-DL-β-phenylalanine, poly-(DL-β-phenylalanine:γ-methyl-L-glutamate), and poly-(DL-β-phenylalanine:L-leucine). Larger steric repulsion of side chains would be expected for these copolymers than for poly-γ-methyl-L-glutamate.

The infrared dichroism observed by Ambrose and Elliott is accounted for very satisfactorily by our helical structure. In the 3.7-residue hydrogen-bonded helix the N—H and C=O bonds are oriented nearly parallel to the helical axis, the angle of deviation being only about 12°. This leads to a predicted dichroic ratio of about 44:1 for the N—H and C=O stretching vibrations. The largest observed dichroism, 14:1 for N—H stretching, is well within the predicted limit; it corresponds to an average angle of deviation of about 20° in the partially oriented specimen.

We conclude that there is strong evidence that crystals of poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate, and also of the peptide copolymers, contain molecules with the third-amide hydrogen-bonded helical configuration, with about 3.7 residues per turn; and that the intermolecular forces operating in the crystals have caused the helixes in the first substance to assume the 18-residue 5-turn configuration, and those in the second to assume the 11-residue 3-turn configuration, these configurations having 3.60 and 3.67 residues per turn, respectively.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The U. S. Public Health Service.

* Contribution No. 1551.

[1] Pauling, L., and Corey, R. B., *J. Am. Chem. Soc.*, **72,** 5349 (1950).

[2] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, 205–211 (1951).

[3] Pauling, L., and Corey, R. B., *Ibid.*, to be published later.

[4] Bamford, C. H., Hanby, W. E., and Happey, F., *Proc. Roy. Soc.*, **A205,** 30 (1951).

[5] Ambrose, E. J., and Elliott, A., *Ibid.*, **A205,** 47 (1951).

[6] Huggins, M. L., *Chem. Rev.*, **32,** 195 (1943).

[7] Zahn, H., *Z. Naturforsch.*, **2B,** 104 (1947).

[8] Simanouti, T., and Mizushima, S., *Bull. Chem. Soc. Japan*, **21,** 1 (1948).

[9] Ambrose, E. J., and Hanby, W. E., *Nature*, **163,** 483 (1949).

[10] Astbury, W. T., *Ibid.*, **164,** 439 (1949).

# THE PLEATED SHEET, A NEW LAYER CONFIGURATION OF POLYPEPTIDE CHAINS

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

For many years it has been assumed that in silk fibroin, stretched hair and muscle, and other proteins with the $\beta$-keratin structure the polypeptide chains are extended to nearly their maximum length, about 3.6 A per residue, and during the last decade it has been assumed also that the chains form lateral hydrogen bonds with adjacent chains, which have the opposite orientation. A hydrogen-bonded layer of this sort is represented diagrammatically in figure 1.[1-4]

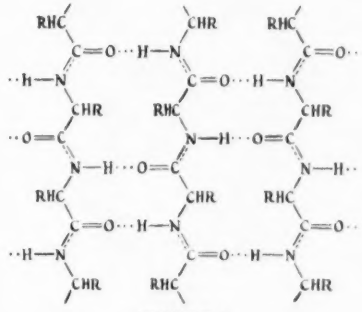


FIGURE 1

Diagrammatic representation of a hydrogen-bonded layer structure of polypeptide chains with alternate chains oppositely oriented.

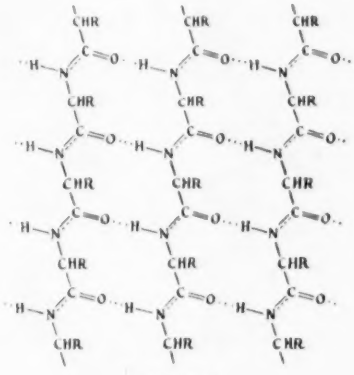


FIGURE 2

Diagrammatic representation of a hydrogen-bonded layer structure of polypeptide chains with all chains similarly oriented (the pleated sheet).

We have now discovered that there is another, rather similar hydrogen-bonded layer configuration of polypeptide chains, which differs from that of figure 1 in several ways. In the new configuration, which we shall call the pleated-sheet configuration, the plane formed by the two chain bonds of the $\alpha$ carbon atom is perpendicular to the plane of the sheet, as shown in figures 2 and 3, rather than being coincident with it. In this structure the successive residues in a chain are similarly oriented, directing their carbonyl groups in one direction and their imino groups in the opposite direction, and all of the chains are oriented in the same way, instead of adjacent chains being opposed in direction.

Let us assume that a polypeptide chain with the configuration indicated diagrammatically in figure 2 is bent in such a way that the planes of the successive amide groups form dihedral angles whose edges are perpendicular to the plane formed by the axes of the groups (the lines connecting successive $\alpha$ carbon atoms). It is found that if the bond distances and bond angles are given the values that we have used in our recent considerations of protein configurations the dihedral angle has the value 106.5°, and the vertical component of the axis of each residue is 3.07 A. It is also found that the carbonyl and imino groups are oriented in such a way that they can form satisfactory hydrogen bonds with corresponding groups in chains

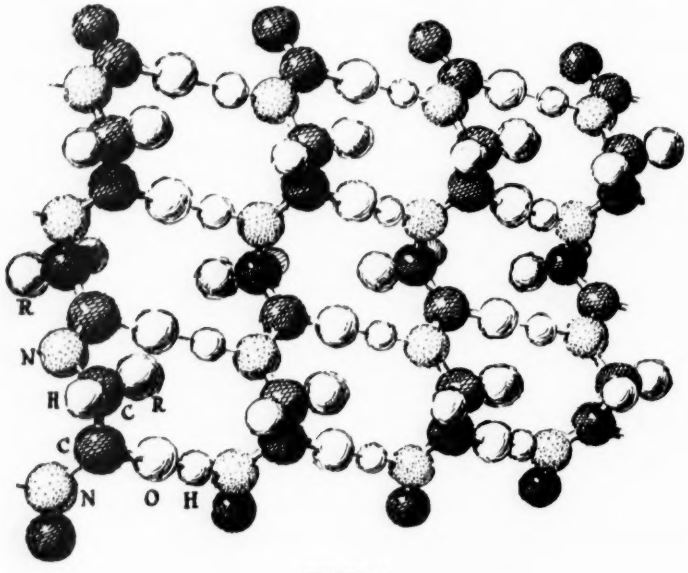



FIGURE 3

Drawing representing the pleated-sheet configuration of polypeptide chains.

obtained by lateral translation. If the lateral translation is given the value 4.75 A the N—H···O distance is 2.75 A; this is a normal hydrogen-bond distance. The N—H axis lies within 6° of the N···O axis, indicating that a stable hydrogen bond should be formed. The coordinates of atoms for the pleated-sheet configuration are given in table 1, and a drawing of the configuration is shown as figure 3.

It is to be noted that each amide group in the chain (neglecting the side chains) may be described as obtained from the preceding one by the operation of a glide plane of symmetry. Because of this, side chains of L-amino

acid residues are related differently to the structure when attached to one $\alpha$ carbon atom than when attached to the $\alpha$ carbon atom of an adjacent residue. The pleated-sheet configuration can accordingly be described as involving only one kind of glycine residue, in case that it were to be assumed by a polyglycine, but two kinds of residues for all optically active amino-acid polymers. These two kinds differ in that, for the L configuration, a residue of one kind points its $\beta$ carbon atom in the C=O direction, and a residue of the other kind points its $\beta$ carbon atom in the N—H direction.

We have found some evidence to support the belief that the pleated-sheet configuration is present in stretched muscle, stretched hair, feather keratin, and some other fibrous proteins that have been assigned the $\beta$-keratin structure. These proteins give x-ray diagrams on which there is a strong meridional reflection corresponding to spacing about 3.3 A, which is a few per cent larger than the fiber-axis distance per residue for the undistorted

TABLE 1

COORDINATES OF ATOMS IN THE POLYPEPTIDE PLEATED-SHEET CONFIGURATION (IN A)

| | UNROTATED | | | 7° ROTATION | | | 20° ROTATION | | |
|---|---|---|---|---|---|---|---|---|---|
| ATOM | $x$ | $y$ | $z$ | $x$ | $y$ | $z$ | $x$ | $y$ | $z$ |
| $C_1$ | 0.00 | 1.15 | 0.00 | 0.00 | 1.09 | 0.00 | 0.00 | 0.96 | 0.00 |
| $N_1$ | −0.36 | 0.30 | 1.14 | −0.36 | 0.46 | 1.17 | −0.36 | 0.35 | 1.29 |
| $C_1'$ | 0.53 | −0.28 | 1.91 | 0.53 | −0.31 | 1.96 | 0.50 | −0.40 | 1.98 |
| $O_1$ | 1.74 | −0.14 | 1.73 | 1.72 | −0.31 | 1.75 | 1.63 | −0.64 | 1.58 |
| $C_2$ | 0.00 | −1.15 | 3.07 | 0.00 | −1.09 | 3.15 | 0.00 | −0.96 | 3.32 |
| $N_2$ | −0.36 | −0.30 | 4.21 | −0.36 | −0.39 | 4.31 | −0.34 | −0.14 | 4.49 |
| $C_2'$ | 0.53 | 0.28 | 4.98 | 0.53 | 0.22 | 5.12 | 0.50 | 0.08 | 5.49 |
| $O_2$ | 1.74 | 0.14 | 4.80 | 1.72 | −0.04 | 4.95 | 1.63 | −0.39 | 5.50 |
| $C_1^*$ | 0.00 | 1.15 | 6.14 | 0.00 | 1.09 | 6.30 | 0.00 | 0.96 | 6.64 |

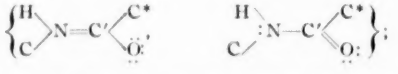
pleated sheet, but much smaller than the value 3.6 A for fully extended polypeptide chains. We have noticed that the pleated sheet can be subjected, without rupturing the hydrogen bonds, to a considerable distortion, in such a way as to increase the fiber-axis distance. This distortion is effected by rotating each amide group about its C—C* axis through a small angle. The rotation moves one of the two $\beta$ positions of each carbon atom farther from the median plane and the other nearer, and the effective rotations for the two non-equivalent kinds of optically active residues are such as to permit each to be an L residue with its side chain farther from the median plane than in the undistorted structure. Presumably the van der Waals repulsion of the side chain atoms and the main chain atoms would be operating in proteins of normal chemical composition with the pleated-sheet configuration, and this would cause some distortion of the chain-lengthening sort. (It is to be noted that two kinds of pleated sheets can be constructed of L-amino-acid residues, of which for one the deformation that

relieves the strain of side chain van der Waals repulsion increases the fiber-axis length, and for the other it decreases it.) It might well occur that the magnitude of the deformation would be such as to give the fiber-axis residue length observed for the $\beta$-keratin proteins, about 3.3 A. This deformation results from a 20° rotation of the amide groups, which gives 3.32 A as the residue length. Coordinates for the structure with 20° rotation and also for a less deformed structure, with 7° rotation, are given in table 1.

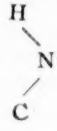
The deformed structures require some distortion of the hydrogen bonds, in that if the hydrogen atom is kept coplanar with the amide group the N—H direction deviates from the N···O axis by an angle somewhat greater than the distorting angle of rotation. The nature of the distortion is such, however, as to suggest that not much strain energy is involved. Let us consider the effect on the stability of the amide group of moving the hydrogen atom onto (or nearly onto) the N···O axis. This motion would keep the hydrogen atom nearly in a plane normal to the plane; that is, it



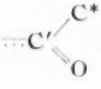
involves moving the hydrogen atom toward one of the tetrahedral corners of the nitrogen atom. If the nitrogen atom were forming a pure double bond with the carbonyl carbon C′ there would be strong resistance to this motion of the hydrogen atom. However, it forms a bond with about one-half double-bond character and one-half single-bond character, corresponding to the resonance



and for the second of the structures the tetrahedral position for the hydrogen atom would be the normal one, whereas for the first the planar position is stable. According we would predict that this rotational distortion of the pleated sheet would not involve so much strain as if the bonds were double bonds.

We may now ask to what extent distortion of the amide group from the planar configuration, through rotation of the two ends  and



 in opposite directions about the N···C′ axis, might be expected

to occur. The strain energy of this distortion, which is essentially also the strain energy of distortion of the hydrogen atom out of the plane, can be estimated in the following way. With $\delta$ the dihedral angle formed by the planes of the two end groups, the amide resonance energy may be taken equal to $-A \sin^2(\delta - \pi/2)$, and the strain energy to $A \sin^2 \delta$. The factor $A$ is the amide resonance energy for the planar configuration. This may be estimated as about 30 kcal mole$^{-1}$. (The experimental value for the carboxylate ion, in which each of the two C···O bonds has 50 per cent double-bond character, is 36 kcal mole$^{-1}$, and somewhat smaller values are found for gas molecules of amides, esters, and related substances.[5]) We thus find about 0.9 kcal mole$^{-1}$ strain energy for 10° distortion of the amide group, 3.5 kcal mole$^{-1}$ for 20° distortion, and so on, and we may predict that distortions as large as 20° might well occur in structures in

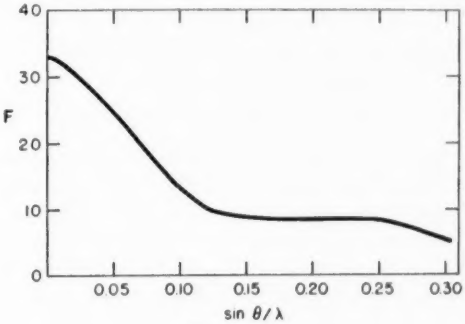



FIGURE 4

Calculated x-ray form factors for the pleated sheet, for planes parallel to the plane of the sheet.

which these distortions would relieve a larger strain, but that in general the polypeptide chain would avoid structures involving such strains. In any case, we would expect the distortion to be divided between the amide residue and the hydrogen bond. In calculating the coordinates of table 1 we have not taken account of these distortions.

The discussion of the pleated sheet in $\beta$ keratin and other proteins will be presented in following papers. In this discussion we make use of the x-ray scattering form factor for the sheet. The form factor, calculated for reflections from planes parallel to the median plane of the undistorted sheet, is for convenient later reference given here, in figure 4, as calculated from the equation $F = \sum_i f_i \cos(2\pi y_i \sin\theta/\lambda)$, with $f_i$ values as given in the International Tables for the Determination of Crystal Structures. The sum has been taken over the atoms of one residue of the undistorted structure, including also a $\beta$ carbon atom, with $y = 2.04$.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The United States Public Health Service.

* Contribution No. 1552.

[1] Astbury, W. T., *Trans. Faraday Soc.*, **29**, 193 (1933).

[2] Huggins, M. L., *J. Org. Chem.*, **1**, 407 (1936); *Chem. Rev.*, **32**, 195 (1943).

[3] Pauling, L., *J. Am. Chem. Soc.*, **62**, 2643 (1940).

[4] Lotmar, W., and Picken, L. E. R., *Helv. Chim. Acta*, **25**, 538 (1942).

[5] Pauling, L., *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, N. Y., 1939.

---

## *THE STRUCTURE OF FEATHER RACHIS KERATIN*

By Linus Pauling and Robert B. Corey

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology, Pasadena, California

Communicated March 31, 1951

The rachis of feathers gives rise to x-ray diffraction patterns of great complexity—they have been described as the most complex known for the naturally occurring fibrous substances. For their interpretation there is required a unit of structure with dimensions at least 9.5 A × 34 A × 94.6 A. In the following paragraphs we propose a structure for this protein that accounts for the principal features of the x-ray pattern and for some physical properties of the substance.

Astbury and other workers in the field have mentioned that the pattern somewhat resembles that of stretched hair, stretched muscle, and other proteins with the $\beta$-keratin structure, but that the x-ray diagram indicates that the length per residue is only 3.07 A, somewhat shorter than expected for an extended polypeptide chain, about 3.6 A, and than observed for silk fibroin, about 3.5 A, and for the $\beta$-keratin proteins, about 3.3 A. Astbury suggested that the chains might be in a somewhat collapsed $\beta$-keratin configuration, and pointed out that the reversible extensibility of feather keratin through about 7 per cent supported this assumption.[1] We were struck by the identity of the indicated fiber-axis residue length, 3.07 A, and the corresponding length predicted for the undistorted pleated-sheet configuration of hydrogen-bonded polypeptide chains, described in the preceding paper, and we investigated the possibility that feather keratin is composed of these pleated sheets in parallel orientation. This can be ruled out as unsatisfactory, however, in that, although the predicted distance between chains in the direction of the hydrogen bonds, 4.75 A, agrees closely with that indicated by the x-ray diagram, about 4.68 A, the other equatorial

reflections on the diagram cannot be accounted for by such a structure. The diagram[3] shows four equatorial reflections, with spacings 34 A, 17 A, 11 A, and 8.56 A, that seem to be the first four orders from a unit with edge about 34 A. The magnitude of this dimension suggests three layers of protein, each about 11 A thick, and these layers could not be identical, inasmuch as all four orders of reflections are observed, rather than only the third order. Consideration of alternative possibilities led us to the conclusion that the layers consist of a pleated sheet and two layers of $\alpha$ helixes, these helixes having the configuration described in recent papers.[4]

A plan of the proposed structure is shown as figure 1, and a schematic drawing as figure 2. The chains of the pleated sheet are at the positions $x = 0$, $y = 0$, and $x = {}^1/_2$, $y = 0$, the base of the unit of structure having the dimensions $a_0 = 9.5$ A and $b_0 = 34.2$ A. The centers of the $\alpha$ helixes

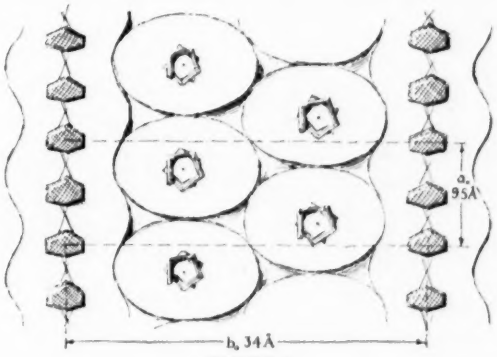



FIGURE 1

Plan of the proposed structure for feather rachis keratin. The structure consists of pleated-sheet layers, between which there are double layers of 3.7-residue helixes.

are at ${}^1/_4$, $y$ and ${}^3/_4$, $\bar{y}$, with $y$ approximately ${}^1/_3$. The $\alpha$ helixes are indicated to be in the close-packed arrangement given by these parameters by the absence of an equatorial reflection at 9.5 A. A diffuse reflection corresponding to the second order is observed at about 4.68 A.

We have found that the intensities of the equatorial reflections can be rather well explained by this structure, with consideration only of the atoms of known fixed position in the pleated sheet and in the $\alpha$ helixes. The atomic parameters and also the form factors for the pleated sheet have been reported in a preceding paper,[2] and those for the $\alpha$ helix are given in the following paper. The value of the parameter $y$ indicated by the data is 0.275. The structure factors calculated with use only of the atoms of known position, including the $\beta$ carbon atom for each residue in both the

pleated sheet and the $\alpha$ helixes, and with the approximation of cylindrical symmetry for the $\alpha$ helixes, are given in table 1 under the heading $F_1$.

Although these structure factors are in rough agreement with the observed intensities, somewhat improved agreement is obtained by making a correction for the remaining atoms of the side chains of the pleated-sheet residues. We have carried out this calculation by assuming two carbon atoms, or atoms of equivalent scattering power, per residue, arranged about the positions $y = +1/12$ and $-1/12$. These positions, at 2.85 A from the center of the pleated sheet, are those expected for the pleated-sheet side chains. The distribution of the atoms about these positions has been approximated by using an $F$-factor for these atoms proportional to that calculated for the pleated sheet.[2] The values $F_2$ in table 1 were calculated with inclusion of these side-chain atoms, and the quantity $I_{calc.}$ given in the following column was obtained by multiplying the square of $F_2$ by the Lor-

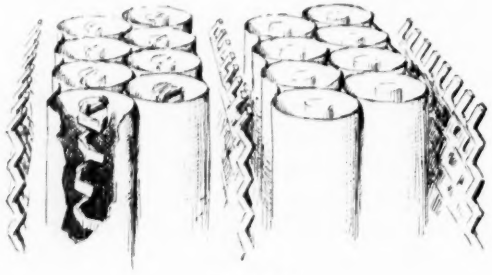


FIGURE 2

Drawing representing the proposed structure for feather rachis keratin.

entz, polarization, and frequency factors. The calculated intensity for (040) seems to be small; however, this calculated intensity is increased to the value 330 by inclusion of the intensity, 175, calculated for the form $\{110\}$, which is the only important diagonal reflection in this region of the equatorial plane. The structure factor for $\{120\}$ is small. The strong, diffuse reflection at 4.68 A is due to $\{200\}$, $\{210\}$, and $\{220\}$, with calculated spacings 4.75 A, 4.71 A, and 4.58 A. The structure factor for the $\alpha$ helix is very nearly 0 for this interplanar distance, and the form factor for these reflections is that of the chain in the $\beta$ sheet, with contributions from the side-chain atoms. The value of $F_2$ has been obtained by assuming that the main-chain and side-chain atoms have the same effective distribution about their central axis as has the pleated sheet in the $y$ direction. The general agreement of observed and calculated intensities is seen to be satisfactory.

In addition to these equatorial reflections, Corey and Wyckoff[3] reported three others, at 51.0 A (faint), 81.8 A (medium), and 115 A (medium). Bear[5] has pointed out that the central regions of the x-ray pattern given by feather rachis are confused by radiation artifacts related to the strong reflections, and that this is probably the source of these three large equatorial spacings. If we accept this interpretation, the equatorial reflections of feather keratin are all accounted for by our unit.

The meridional reflections observed by Corey and Wyckoff[3] and by Bear[5] are given in table 2. They can nearly all be accounted for as orders of a large identity distance, 94.6 A. In addition to the meridional reflections given in table 2, Bear reported a number of small-angle near-meridional reflections, corresponding to other orders of the identity distance 94.6 A.

TABLE 1

COMPARISON OF CALCULATED AND OBSERVED INTENSITIES OF EQUATORIAL X-RAY REFLECTIONS FOR FEATHER RACHIS KERATIN

PSEUDO-ORTHORHOMBIC, TRICLINIC UNIT WITH $a_0 = 9.50$ A, $b_0 = 34.2$ A, $c_0 = 94.6$ A, $\alpha \cong 90°$, $\beta \cong 90°$, $\gamma \cong 90°$

| *hkl* | $d_{calc.}$ | $F_1$ | $F_2$ | $I_{calc.}$ | $I_{obs.}$[a] | $d_{obs.}$[a] |
|---|---|---|---|---|---|---|
| 010 | (34.2 A) | 22.8 | 42.6 | 730 | Strong | 33.3 A |
| 020 | 17.1 | −20.8 | −10.4 | 24 | Faint | 17.1 |
| 030 | 11.4 | 45.5 | 45.5 | 295 | Medium | 11.0 |
| 110 | 9.14 | −30.0 | −30.0 | 210 | Medium | 8.56 |
| 040 | 8.56 | 46.1 | 37.6 | 155 | | |
| 050 | 6.84 | 9.4 | −3.2 | 1 | ... | ... |
| 060 | 5.70 | 12.3 | 0.4 | 0 | ... | ... |
| 070 | 4.89 | 11.4 | 3.1 | 1 | ... | ... |
| 200 | 4.75 | ... | 25 | 38 | Strong (diffuse) | 4.68 |
| 210 | 4.71 | ... | 25 | 75 | | |
| 220 | 4.58 | ... | 25 | 75 | | |
| 080 | 4.28 | 10.6 | 6.6 | 2 | | |
| 090 | 3.80 | 16.4 | 16.4 | 12 | | |
| 0·10·0 | 3.42 | 9.0 | 12.2 | 6 | | |

[a] Corey and Wyckoff.[3]

For the ideal pleated sheet we have calculated a fiber-axis distance of about 3.07 A per residue, which agrees in numerical value with the spacing of the outermost meridional reflection. We think it likely, however, that the fiber-axis length per pleat in the pleated sheet in this substance is in fact not 6.14 A, but 6.30 A, and that there are fifteen of these units (thirty amino-acid residues) in the unit with $c_0 = 94.6$ A. The corresponding fiber-axis distance per residue, 3.15 A, can be achieved by a small deformation of the ideal pleated sheet, amounting to rotation by 7° around the C—C* axes of the amide groups, as described in the preceding paper.[2] A possibility that seems to us less likely is that there are $15\frac{1}{2}$ units of the pleated sheet in the distance 94.6 A, in which case the length per residue would be 3.05 A, and the value of $c_0$ would be twice as great, 189.2 A.

The fiber-axis length per residue for the $\alpha$ helix in substances so far investigated ranges from 1.53 to 1.56 A. These values correspond to the integers 62 and 61, respectively, as the number of residues in unit length 94.6 A along the $c$ axis, the value 62 giving 1.525 A and 61 giving 1.551 A as the length per residue. The number of residues per unit turn is predicted to be close to 3.69, and values between 3.6 and 3.67 have so far been reported. With 17 turns in $c_0$, 61 residues would give 3.59 residues per turn, and 62 residues would give 3.65 residues per turn; no other value than 17 turns seems likely, inasmuch as it is improbable that the $\alpha$-carbon bond angle would be strained enough to give 3.81 residues per turn (61 residues in 16 turns) or 3.45 residues per turn (62 residues in 18 turns). The most likely possibility is thus the 62-residue 17-turn helix, with 3.65 residues per turn. The distribution of side chains along the polypeptide chain presum-

TABLE 2

MERIDONAL REFLECTIONS FROM FEATHER RACHIS KERATIN

| ORDER OF REFLECTION | BEAR | | COREY AND WYCKOFF | |
|---|---|---|---|---|
| | $d$ | INTENSITY | $d$ | INTENSITY |
| 4 | 23.6 A | 10 | 23.1 A | Strong |
| | | | 17.2 | ? |
| 8 | 11.90 | 3 | ... | ... |
| 9 | 10.46 | 3 | ... | ... |
| ? | ... | .. | 9.08 | Faint |
| 15 | 6.30 | 6 | 6.20 | Strong |
| 17 | 5.53 | 3 | ... | ... |
| 19 | 4.98 | 6 | 4.90 | Strong |
| 21 | 4.45 | 4 | 4.37 | Medium |
| 24 | ... | .. | 3.95 | Faint |
| 27 | ... | .. | 3.52 | Faint |
| ? | ... | .. | 3.22 | Faint |
| 31 | ... | .. | 3.07 | Medium |

ably is such as to stabilize this helix, and the scattering of x-rays by the different side chains, in positions that remain to be determined, gives rise to the true meridional reflections, with contributions from the atoms of the main chains in the helix and the pleated sheet for some reflections.

It is interesting to note that the reflection (0·0·17) is reported by Bear, at 5.35 A; this would correspond to one turn of the $\alpha$ helix. A strong reflection at 6.20 A (Corey and Wyckoff) or 6.30 A (Bear) is interpreted by Bear as the fifteenth basal plane reflection, (0·0·15). We interpret this as involving collaboration of the 15 units of the pleated sheet.

The identity distance 94.6 A receives a rational explanation as resulting from the presence of two structures, the slightly distorted pleated sheet with identity distance along the $a$ axis of 6.30 A, and the $\alpha$ helix with 5.57 A per turn. The simplest ratio of integers approximating the ratio of these numbers is 17:15, corresponding to the mutual identity distance

94.6 A. These two structures accordingly might well be expected to form a protein such as feather keratin, with a triclinic unit with $a_0 = 9.50$ A, $b_0 = 34.2$ A, $c_0 = 94.6$ A, $\alpha \cong 90°$, $\beta \cong 90°$, $\gamma \cong 90°$.

It seems likely that the pleated sheets are all oriented similarly in the structure—there is no significant indication of a unit with $b_0 = 68$ A, corresponding to two kinds of pleated sheets, with opposite orientations. A pleated sheet is polar: all of the C=O groups point in one direction, and the N—H groups in the opposite direction, and in addition the side chains on one side of the sheet are arranged differently with respect to the residues than are those on the other side of the sheet, so that an isolated sheet would be curved. It is interesting to speculate that this curvature of the pleated sheets may be related to the natural curvature of the feather rachis. An example of a polar sheet in the inorganic field is the kaolin sheet.[6] Curved crystals of the clay minerals have been recently observed with use of the electron microscope, and their curvature has been assumed to result from the polar nature of the kaolin sheets.[7, 8]

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The United States Public Health Service.

* Contribution No. 1553.

[1] Astbury, W. T., Cold Spring Harbor Symposia on Quantitative Biology, **2,** 15 (1934).

[2] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 251 (1951).

[3] Corey, R. B., and Wyckoff, R. W. G., *J. Biol. Chem.*, **114,** 407 (1936).

[4] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37,** 205 (1951); Pauling, L., and Corey, R. B., *Ibid.* **37,** 235 (1951).

[5] Bear, R. S., *J. Am. Chem. Soc.*, **66,** 2043 (1944).

[6] Pauling, L., these PROCEEDINGS, **16,** 123 (1930).

[7] Davis, D. W., Rochow, T. G., Rowe, F. G., Fuller, M. L., Kerr, T. S., and Hamilton, T. K., Electron Micrographs of Reference Clay Minerals, Preliminary Report No. 6 of American Petroleum Institute Project 49 (1950).

[8] Bates, T. F., Hildebrand, S. A., and Swineford, A., Morphology and Structure of Endelite and Halloysite, *American Mineralogist*, **35,** 463 (1950).

---

## *THE STRUCTURE OF HAIR, MUSCLE, AND RELATED PROTEINS*

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

It is thirty years since x-ray photographs were first made of hair, muscle, nerve, and sinew, by Herzog and Jancke.[1] During this period, despite the efforts of many investigators, the photographs have eluded detailed in-

terpretation, and the molecular structures of the proteins have remained undetermined.* In the present paper we propose structures for hair, muscle, and related proteins in the extended state ($\beta$ keratin and $\beta$ myosin) and the contracted state ($\alpha$ keratin and $\alpha$ myosin), and discuss the extent to which the diffraction data are accounted for by the proposed structures.

*The α-Keratin Structure.*—It seems not unlikely that the polypeptide chains in unstretched hair, contracted muscle, horn, nail, quill, and other proteins that give the $\alpha$-keratin x-ray pattern have the 3.7-residue helical configuration[2] (which for convenience we shall call the $\alpha$ helix).

Let us consider the structure expected for an aggregate of $\alpha$ helixes. The molecules, with the approximate form of circular cylinders, would be expected to pack in a hexagonal or pseudohexagonal array, as do the synthetic polypeptides poly-γ-methyl-L-glutamate and poly-γ-benzyl-L-glutamate.[3] The average residue weight 120 and approximate density 1.30 g cm$^{-3}$ lead to 11 A as the value of $a_0$ for the hexagonal unit. The predicted equatorial reflections are $\{10\cdot0\}$ at 9.5 A, $\{11\cdot0\}$ at 5.5 A, $\{20\cdot0\}$ at 4.8 A, $\{21\cdot0\}$ at 3.6 A, $\{30\cdot0\}$ at 3.2 A, etc. The observed pattern of $\alpha$ keratin, as described by Astbury and Street,[4] has a strong equatorial reflection at 27 A, a very strong reflection at about 9.8 A, and a vague region of darkening around 3.5 A. We would attribute the 27-A reflection to a long-range order that can be elucidated only through further study. The 9.8-A reflection is described as covering a range of spacings of about 3 A centered at 9.8 A, which suggests that the packing is only pseudo-hexagonal, and that the hexagonal form $\{10\cdot0\}$ is split into several forms with different spacings. The failure to observe the reflections $\{11\cdot0\}$ and $\{20\cdot0\}$ can be attributed to the smallness of the x-ray form factor for equatorial scattering by the $\alpha$ helix,[5] which has a node at 5 A; indeed, the $\alpha$-keratin x-ray photographs show a light band which is centered at about 5 A. The form factor then has a maximum at 3.4 A, which corresponds to the vague region of darkening, with center around 3.5 A, reported by Astbury and Street.

The principal meridional feature of the $\alpha$-keratin x-ray pattern is a strong arc at 5.15 A. This reflection has been accepted as indicating that the $c$-axis identity distance is 5.15 A or a simple multiple of it, and it has usually been assumed that the $c$-axis length per residue is either $^1/_3\cdot5\cdot15 = 1.72$ A or $^1/_2\cdot5\cdot15 = 2.58$ A. The 5.15-A arc seems on first consideration to rule out the $\alpha$ helix, for which the $c$-axis period must be a multiple of the axis distance per turn, which is about 5.6 A. However, it was noted by Bamford, Hanby, and Happey[6] that the "meridional arc" observed on photographs of partially oriented fibers of poly-γ-benzyl-L-glutamate is on other photographs resolved into off-meridional spots in positions corresponding to the value 5.76 A for the $c$-axis translation. It seems probable that the 5.15-A arc seen on the $\alpha$-keratin photographs is to be interpreted in a similar way.

A very significant contribution has been made by Herzog and Jancke[7] in 1926 and Lotmar and Picken[8] in 1942. These investigators obtained, by non-reproducible procedures, preparations of rather well crystallized muscle. Herzog and Jancke observed eight forms, and Lotmar and Picken eighteen. Lotmar and Picken's preparation was a piece of posterior-valve-closing muscle (of the mussel *Mytilus edulis*) that had been dried for 48 hours under 10-g tension and then allowed to stand in a can for a year. They indexed their photograph and (also Herzog and Jancke's) with a monoclinic unit with $a_0 = 11.70$ A, $b_0 = 5.65$ A (fiber axis), $c_0 = 9.85$ A, and $\beta = 70.5°$. The fiber-axis translation 5.65 A is in fine accordance with prediction for the 3.7-residue helix.

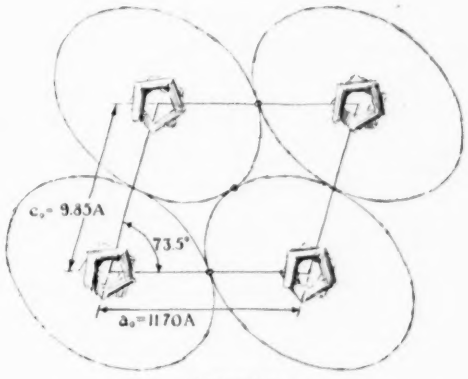



FIGURE 1

Plan of the monoclinic unit proposed for crystalline muscle by Lotmar and Picken, with the 3.7-residue helixes, suggested in the present paper, represented by elliptical cylinders. The dots indicate the positions assumed for side-chain carbon atoms.

Lotmar and Picken stated that their excellent photograph presumably shows the x-ray diagram of crystallized myosin, and that there is no sign of the 5.15-A meridional arc. They considered their preparation to represent a new molecular modification of myosin, with two residues per 5.65-A length along the *c*-axis. We think it likely that the process of crystallization has involved only the ordering of the $\alpha$ helixes, as shown in figures 1 and 2, in such a way that the larger side chains are grouped into layers with $x \cong 1/2$, the region near $x = 0$, $z = 1/2$ being free of side chains, thus permitting the helixes to come into close contact at these points.

The occurrence of the 5.15-A arc on the x-ray photographs of poorly ordered aggregates of $\alpha$-helical molecules can be explained by the consid-

eration illustrated in figure 3. Here a helical curve with pitch 5.65 A is shown on a cylinder with radius 1.81 A, the average radius for the peptide atoms C, N, C′, and O. The angle of inclination of the helix is 26°, and the perpendicular distance between adjacent turns of the curve is 5.65 cos 26° = 5.1 A. We would thus predict strong reenforcement of x-rays scattered by an $\alpha$ molecule in directions about 26° from the fiber axis. (A related pertinent fact is that the maximum for the calculated radial distribution function for the $\alpha$ helix comes at 5.0 A; this function will be reproduced in a later paper.) In the case of a somewhat poorly ordered fibrous aggregate of the molecules or of a crystalline phase with very large unit the intermolecular interference would in general permit strong diffraction maxima with spacing 5.1 A to occur in the near-meridional region, whereas

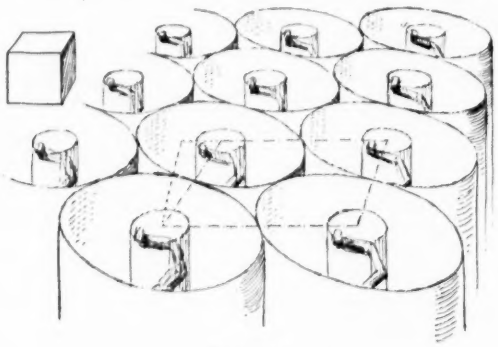


FIGURE 2

Drawing of the proposed structure of $\alpha$ keratin.

for a well-crystallized specimen intermolecular interference could cause these reflections to fail to appear, despite their large molecular form factor.

(Added April 10, 1951: It has been pointed out to us by Professor Verner Schomaker that the foregoing argument is not reliable. He has evaluated the form factor for x-ray scattering by a uniform helix, and has found that the maximum scattering by the helix with the dimensions given above occurs at angles considerably larger than 26° from the meridional direction, and at a Bragg angle corresponding to a spacing of about 4.2 A, rather than 5.1 A. He has further shown that if four main-chain atoms and the $\beta$ carbon atom are represented by helixes with the correct radii for the 3.7-residue helical structure, and four side-chain atoms are represented by another helix, with the same pitch and with radius 4.0 A, a pronounced maximum is predicted to occur at 26° from the meridional direction and at a Bragg angle corresponding to a spacing of 5.1 A.)

The interplanar distances and estimated intensities of the basal-plane reflections on the x-ray photographs of crystalline muscle reported by Lotmar and Picken and by Herzog and Jancke are given in table 1, together with calculated values of the interplanar distance $d$, structure factor $F$, and intensity $I = LPF^2$, with $L$ the Lorentz factor and $P$ the polarization factor. The structure factor is that for the $\alpha$ helix, including a $\beta$ carbon atom for each residue, evaluated for the case of cylindrical symmetry,[5] plus a side-chain carbon atom per residue at $x = {}^1/_2$, $z = 0$ and another one at $x = {}^1/_2$, $z = {}^1/_2$. If the crystal has monoclinic symmetry, as assumed by Lotmar and Picken, there are 2-fold screw axes passing through these points, and a greater-than-average density of atoms might be expected near these axes; we have accordingly tried to approximate the effect of the side-chain atoms on the structure factor by placing atoms in these positions.

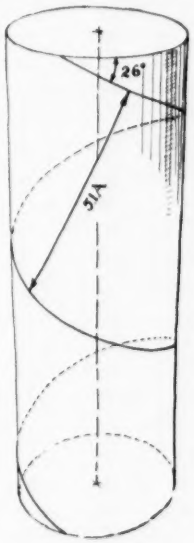



FIGURE 3

Diagrammatic representation of the 3.7-residue helix, indicating the origin of the meridional arc at 5.15 A observed on x-ray photographs.

It is seen that the calculated intensity pattern corresponds surprisingly well with the observed pattern, when it is considered that no variable parameter is involved in the calculation. The only arbitrary decision involved in the calculation is that the side-chain scattering, for about four heavy atoms per residue, can be approximated by placing a carbon atom near each of two 2-fold axes.

The first reflection, $A_1$, is seen on the reproduction of their photograph in Lotmar and Picken's paper to be very strong—we estimate it to be perhaps ten times as strong as $A_2$ or $A_4$. Its breadth is about 3 A, which is enough to include the forms $\{100\}$, $\{001\}$, and $\{101\}$. This spot closely resembles the corresponding spot given by ordinary preparations of $\alpha$ keratin, and described by Astbury and Street as covering a range of about 3 A, centered at 9.8 A. We accordingly think that it is likely that the ordinary preparations of $\alpha$ keratin have in fact a monoclinic structure closely resembling that of crystalline muscle, and only rather slightly disordered. The 27-A reflection seen on photographs of hair and ordinary muscle seems not to be present on Lotmar and Picken's photograph.

The second reflection, $A_2$, is observed to be strong, and calculated as weak. An additional side-chain carbon atom in phase for this reflection (near the line $x + z = 1$) would bring the value of $I_{calc.}$ to 103. (We have not attempted to find a distribution of side-chain atoms that would give

TABLE 1

COMPARISON OF CALCULATED AND OBSERVED INTENSITIES OF EQUATORIAL REFLECTIONS FOR CRYSTALLINE MUSCLE

TRICLINIC UNIT WITH $a_0 = 11.70$ A, $b_0 = 5.65$ A, $c_0 = 9.85$ A, $\alpha \cong 90°$, $\beta = 73.5°$, $\gamma \cong 90°$

| REFLECTION | $hkl$ | $d_{calc.}$ | $F_{calc.}$ | $I_{calc.}$ | $I_{obs.}$ LP[a] | $I_{obs.}$ HJ[a] | $d_{obs.}$ LP[a] | $d_{obs.}$ HJ[a] |
|---|---|---|---|---|---|---|---|---|
| $A_1$ | 100 | 11.21 A | 10.0 | 150 | | | | |
| | 001 | 9.44 | 16.8 | 353 | S (broad) | S (broad) | 9.3 A | 10.0 A |
| | 101 | 8.51 | 14.0 | 217 | | | | |
| $A_2$ | 101 | 6.38 | 6.1 | 32 | S | M | 6.6 | 6.4 |
| $A_3$ | 201 | 5.56 | 2.1 | 3 | ?[b] | M | 6.0 | 5.8 |
| | 200 | 5.10 | 9.7 | 52 | | | | |
| $A_4$ | 102 | 4.87 | −9.8 | 60 | S | ... | 4.78 | ... |
| | 002 | 4.72 | 7.7 | 35 | | | | |
| | 201 | 4.31 | −3.1 | 5 | | | | |
| | 202 | 4.25 | 5.3 | 15 | | | | |
| $A_5$ | 102 | 3.97 | −12.1 | 73 | W | W | 3.97 | 3.94 |
| | 301 | 3.87 | −4.3 | 9 | | | | |
| $A_6$ | 300 | 3.74 | −12.7 | 74 | M | ... | 3.73 | ... |
| $A_7$ | 302 | 3.44 | −12.4 | 65 | M | ... | 3.50 | ... |
| | 103 | 3.28 | −5.0 | 10 | | | | |
| | $20\bar{2}$ | 3.19 | 2.1 | 2 | | | | |
| | $30\bar{1}$ | 3.18 | −4.9 | 9 | | | | |
| | 003 | 3.15 | −4.9 | 9 | | | | |
| | 203 | 3.15 | −4.9 | 9 | | | | |
| $A_8$ | 401 | 2.92 | −4.4 | 7 | ?[b] | ... | 2.94 | |
| | 303 | 2.84 | −4.2 | 6 | | | | |
| | 103 | 2.83 | −4.2 | 6 | | | | |
| | 400 | 2.80 | 2.9 | 3 | | | | |
| $A_9$ | 402 | 2.78 | 2.6 | 2 | W | ... | 2.75 | |
| | 302 | 2.59 | −9.6 | 28 | | | | |
| | 401 | 2.50 | −3.0 | 3 | | | | |
| | 403 | 2.47 | −2.9 | 2 | | | | |
| | $20\bar{3}$ | 2.46 | −2.9 | 2 | | | | |
| $A_{10}$ | 104 | 2.45 | −8.8 | 22 | W | ... | 2.48 | |
| | 204 | 2.44 | 3.1 | 3 | | | | |
| | 004 | 2.36 | 3.2 | 3 | | | | |
| | 501 | 2.34 | −2.4 | 2 | | | | |
| | 304 | 2.31 | −7.8 | 16 | | | | |
| | 502 | 2.29 | −7.7 | 15 | M | ... | 2.18 | |
| $A_{11}$ | 500 | 2.24 | −7.3 | 13 | | | | |
| | 104 | 2.19 | −6.9 | 11 | | | | |
| | $40\bar{2}$ | 2.16 | 3.8 | 5 | | | | |
| | 503 | 2.14 | −1.3 | 0.4 | | | | |
| | $30\bar{3}$ | 2.13 | −1.3 | 0.4 | | | | |
| | 404 | 2.13 | 3.9 | 4 | | | | |

[a] LP = Lotmar and Picken, HJ = Herzog and Janeke.

[b] Presumably the symbol means that Lotmar and Picken were doubtful as to the presence of these two spots.

better general agreement with the entire observed pattern, inasmuch as the approximate agreement given by the less arbitrary calculation that we have made seems to us to be more significant.)

The questionable reflections $A_3$ and $A_8$ cannot be seen on the reproduced photograph. $A_5$, which is described by Lotmar and Picken as weak (whereas $I_{calc.}$ is as large for it as for $A_6$ and $A_7$, described as strong), lies in a region of general blackening, which may have caused its intensity to be underestimated—it is interesting that Herzog and Jancke report $A_5$, but not $A_6$ nor $A_7$. Whether $A_9$ can be assigned to the form $\{30\bar{2}\}$ is doubtful; but in any case the appearance of the three reflections $A_9$, $A_{10}$, and $A_{11}$ in nearly the calculated positions and with the calculated intensities can hardly be the result of coincidence. The sequence of forms with very small values of $I_{calc.}$ in the regions for which no reflections are reported by Lotmar and Picken and also the rather striking agreement found for the observed reflections strongly favor the conclusion that the assumed structure is not greatly different from the actual structure.

Some evidence for the 3.7-residue helix is provided also by the meridional reflections reported for muscle fibers and porcupine quill by Corey and Wyckoff,[9] MacArthur,[10] and Bear.[11] These reflections correspond to large $c$-axis identity distances, about 726 A for muscle fiber and 198 A for porcupine quill. The 726-A unit for muscle is shown also in electron micrographs of muscle fibrils treated with osmic acid.[12] We would expect that side chains of different kinds on the $\alpha$ helix would repeat after an integral number of residues, corresponding to an integral multiple of the residue length along the helix axis, about 1.53 A, and that accordingly those orders of basal plane reflection for which the spacing approximated closely to certain multiples of 1.53 A would be enhanced in intensity. It is in fact found that about 80 per cent of the meridional reflections are of this type; for both Venus clam muscle and porcupine quill they are multiples of 1.51 A. The reflections at 1.49 A, 3.05 A, 4.50 A, and 6.19 A for porcupine quill, which represent the first four orders of enhancement, are the strongest features of the wide-angle meridional pattern, except for the 5.2-A arc.

*The β-Keratin Structure.*—Hair and muscle can be reversibly stretched to about 100 per cent elongation.[13] Some authors have expressed doubt as to whether this elongation is to be attributed to the polypeptide chains, but it seems to us that Astbury's contention that it should be is justified. With a fiber-axis length of 1.53 A per residue for the $\alpha$ helix, an extended chain in the $\beta$-keratin structure would be predicted, on this assumption, to have a fiber-axis residue length of about 3.1 A. The principal meridional x-ray reflection of stretched hair, stretched muscle, and other proteins with the $\beta$-keratin structure[4] has in fact a spacing reported by Astbury as about 3.32 A, which is presumably the fiber-axis residue length, and would thus correspond to 117 per cent extension of the $\alpha$ helix. That the $\beta$-kera-

tin structure involves extended polypeptide chains was first suggested by Brill[14] in 1923, and for the past fifteen years it seems to have been rather generally assumed that the chains are essentially coplanar, and that they alternate in direction in forming hydrogen-bonded non-polar sheets.[15–17, 8] Another hydrogen-bonded layer structure, the pleated sheet, which avoids the difficulty of large steric interference of side-chain groups predicted for the planar sheet, has recently been described.[18] The pleated sheet can easily assume a configuration corresponding to the residue-length 3.32 A, and it seems to us likely that it represents the $\beta$-keratin structure.

At present there is not much direct evidence to support this view. The observed equatorial x-ray reflections[4] at 9.8 A (strong) and 4.65 A (very strong) have been shown by Astbury and Sisson[19] to correspond to the plane of the $\beta$-keratin layers and the lateral direction in the layers, respectively. The lack of other equatorial reflections, except one weak reflection at 2.4 A, and lack of knowledge of positions of side-chain atoms make a calculation of intensities of reflection of little value. It may be pointed out, however, that if the one-molecule unit that accounts for the x-ray pattern of crystalline muscle obtained by Lotmar and Picken is correct the $\alpha$ helixes must all be oriented in the same sense, and accordingly this muscle on stretching could be transformed into the pleated sheet, but not into the planar sheet.

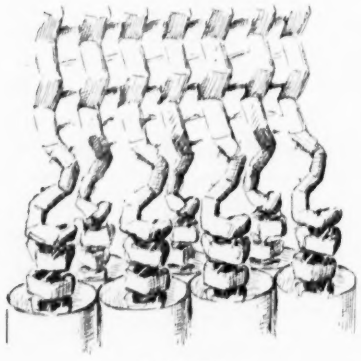


FIGURE 4

Drawing illustrating the proposed mechanism of conversion of a pleated sheet into a double row of 3.7-residue helixes; this is proposed as the process involved in the contraction of muscle.

When hair or muscle is treated with hot water or steam it shortens in the direction of the fiber axis, and swells laterally. The resultant material is called supercontracted keratin. The contraction from the $\alpha$ state is about 30 per cent for both hair[20] and myosin.[21] It is possible that supercontracted keratin has the configuration of the 5.1-residue helix,[2, 5] but there is very little evidence to support this suggestion. The fiber-axis residue length for this helix is about 0.99 A, which corresponds to 35 per cent contraction from the $\alpha$ helix, with residue length 1.53 A. The agreement of this value with the experimental value for the amount of supercontraction provides some support for the suggestion that the $\gamma$ helix is present in supercontracted keratin. A careful study of the x-ray diagram should permit a decision on this point to be made.

*The Mechanism of Contraction of Muscle.*—The assignment of the pleated-sheet configuration to extended muscle and of the $\alpha$-helical configuration to contracted muscle suggests that a discussion be made of the mechanism of contraction of muscle.

We have noticed that in order for a pleated sheet to be converted into a double layer of $\alpha$ helixes it is not necessary that all of the hydrogen bonds in the pleated sheet be initially broken. Instead, it is necessary to break only enough hydrogen bonds, four or five, to liberate four or five residues in each polypeptide chain, these being at about the same horizontal level in the pleated sheet. The liberated chains can then coil into the $\alpha$-helical configuration, to produce the double layer of packed cylindrical $\alpha$ helixes, as shown in figure 4.

The sheet configuration is, for a normal protein, somewhat unstable relative to the $\alpha$ helix. This is indicated to be the case for synthetic polypeptides involving residues other than glycine by the fact that these polypeptides have been observed to form crystals of the $\alpha$ type.[6] Indication that the instability of the pleated sheet is due to steric repulsion between side chains is given by the fact that copolymers containing a large fraction of glycine residues assume the $\beta$ configuration.[6] The steric-hindrance explanation of the instability of the pleated sheet is made reasonable by a consideration of the area available per side chain. In a normal $\beta$ keratin, with fiber-axis length per residue 3.32 A and side-chain spacing 4.75 A and both sides of the sheet available for side chains, the area per side chain is 31.6 $A^2$. For the $\alpha$ helix we may take the radius of the side-chain median cylindrical surface to be 4.16 A, which is midway between the centers of the $\beta$ carbon atoms (radius 3.34 A) and the point of contact with adjacent molecules ($^1/_2c_0$ = 4.98 A); with the fiber-axis residue length 1.53 A, the area per side chain is calculated to be 40.0 $A^2$, which is 25 per cent larger than for the pleated sheet. Moreover, in the discussion of the pleated sheet we have pointed out that the ideal pleated sheet has fiber-axis length per residue 3.07 A, and that the extension to 3.32 A, as observed in myosin, seems to be associated with a steric interference with the side chains, which causes the residues to rotate out of the position most favorable to hydrogen-bond formation, and thus introduces a strain, essentially of bending, in the hydrogen bonds. The normal hydrogen-bond energy for peptides may be estimated at 7.5 kcal $mole^{-1}$; it is expected to be large, because of the negative charge conferred on the carbonyl oxygen atom and the positive charge conferred on the imino nitrogen atom by the amide resonance. The energy of the strain introduced may be estimated to be of the order of magnitude of about 3 per cent of the hydrogen-bond energy, or 0.22 kcal $mole^{-1}$, which is about 1.8 cal per gram of myosin. Since muscle is about 10 per cent myosin, we estimate (very roughly) the work that could be done by 1 g of muscle to be about 0.18 cal, in a single complete

twitch. Muscle has density 1.06 g $cm^{-3}$, and 1 g of muscle in the extended state with cross-section 1 $cm^2$ would be 0.94 cm long. If the entire muscle were to shorten proportionately to the myosin molecules, and these molecules were to shorten from the residue length 3.32 A (for the pleated sheet) to 1.53 A (for the $\alpha$ helix), the contracted length would be 0.43 cm, the contraction being by 54 per cent. The contracting pleated sheet would be predicted to exercise the same force throughout its contraction; for initial cross-section 1 $cm^2$ (in the extended state) this force is calculated from the assumed strain energy 0.18 cal per g of muscle, assuming it to be free energy, to have the value 1.53 kg.

The foregoing rough calculation agrees well with experiment in some respects. A. V. Hill has reported[22] that the maximum force exerted in a twitch by frog's muscle at 0°C is 1 to 2 kg $cm^{-2}$, which agrees well with the value 1.5 kg $cm^{-2}$ calculated above. It is interesting also that the observed maximum shortening,[23] by 50 to 60 per cent for toad muscle, is close to the predicted shortening, 54 per cent, for the transition from the pleated sheet to the $\alpha$ helix. The heat liberated by frog's sartorius muscle on tetanic contraction has been measured,[24] and found to be about 400 g cm per cm of shortening and per $cm^2$ cross-section, or 0.05 cal per g of muscle, assuming complete contraction (by 54 per cent). Somewhat smaller values were found in a later study.[23] These values are considerably smaller than the estimated energy difference of the extended and the contracted forms of the myosin of muscle, as given above (0.18 cal $g^{-1}$).

In order to account for the observed mechanical properties of hair (great extension under a load of 500 to 2000 kg $cm^{-2}$, depending on humidity) in the same way, it must be assumed that the strain of the $\beta$ configuration is very much greater, about 10 cal $mole^{-1}$ per residue. Moreover, the great dependence on humidity shows that side-chain interactions, which are changed by hydration, are involved.

The pleated-sheet configuration of extended muscle is metastable. In order for the polypeptide chains to contract an excitation energy would be needed, the energy of breaking four or five hydrogen bonds, to liberate four or five residues, in each chain. We suggest that the mechanism whereby the reaction of contraction is initiated may involve the production in or transfer to this region of the muscle of a number of hydrogen-bond-forming molecules, which can attack the hydrogen bonds of the pleated sheet, and through the formation of hydrogen bonds with the carbonyl and imino groups of the chains decrease the energy of activation of the sheet-disrupting process. As the $\alpha$ helixes are formed these molecules are liberated, and might continue to attack hydrogen bonds in the pleated sheet. It is, however, not necessary that they do so, in order for the process of formation of $\alpha$ helixes to continue, once that it is started: as the freed residues coil into the $\alpha$ helixes, and form more stable hydrogen bonds within these helixes, they would exercise a mechanical strain, communicated along the

polypeptide chain, on the adjacent residues that are still held by hydrogen bonds in the pleated sheet, and this strain would have the effect of reducing the activation energy for the liberation of further residues. Accordingly once that the reaction were initiated, it would be expected to continue until all of the pleated sheet had been converted into a double row of $\alpha$ helixes.

It is not so easy to suggest a single reasonable way in which the muscle can be reconverted to the stretched state. There are many conceivable ways in which this could be done, with the use of chemical reactions of various sorts—especially the change in the nature of the environment of the $\alpha$ helixes. One possibility, suggested over twenty years ago by Meyer and Mark,[25] is that through a change in the ionic environment in the muscle the polypeptide chain is provided with a sequence of similarly charged side chains, not paired with neutralizing charges of the opposite sign. The electrostatic repulsion of these side chains would then tend to cause the chains to stretch out into the pleated-sheet configuration. It is of interest that Meyer and Mark illustrated this mechanism with use of a simple helical curve for the molecule of contracted muscle.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The United States Public Health Service.

* Contribution No. 1554.

[1] Herzog, R. O., and Jancke, W., "Festschrift der Kaiser Wilhelm Gesellschaft," 1921.

[2] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37,** 205 (1951).

[3] Pauling, L., and Corey, R. B., *Ibid.*, **37,** 241 (1951).

[4] Astbury, W. T., and Street, A., *Phil. Trans. Roy. Soc.*, **A230,** 75 (1931).

[5] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 235 (1951).

[6] Bamford, C. H., Hanby, W. E., and Happey, F., *Proc. Roy. Soc.*, **A205,** 30 (1951).

[7] Herzog, R. O., and Jancke, W., *Naturwiss.*, **14,** 1223 (1926).

[8] Lotmar, W., and Picken, L. E. R., *Helv. Chim. Acta*, **25,** 538 (1942).

[9] Corey, R. B., and Wyckoff, R. W. G., *J. Biol. Chem.*, **114,** 407 (1936).

[10] MacArthur, I., *Nature*, **152,** 38 (1943).

[11] Bear, R. S., *J. Am. Chem. Soc.*, **66,** 2043 (1944).

[12] Jakus, M. A., Hall, C. E., and Schmitt, F. O., *Ibid.*, **66,** 313 (1944).

[13] Astbury, W. T., and Woods, H. J., *Nature*, **126,** 913 (1930).

[14] Brill, R., *Ann. d. Chem.*, **434,** 204 (1923).

[15] Astbury, W. T., *Trans. Faraday Soc.*, **29,** 193 (1933).

[16] Huggins, M. L., *J. Org. Chem.*, **1,** 407 (1936).

[17] Pauling, L., *J. Am. Chem. Soc.*, **62,** 2643 (1940).

[18] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 251 (1951).

[19] Astbury, W. T., and Sisson, W. A., *Proc. Roy. Soc.*, **A150,** 533 (1935).

[20] Astbury, W. T., and Woods, H. J., *Phil. Trans. Roy. Soc.*, **A232,** 333 (1933).

[21] Astbury, W. T., and Dickinson, S., *Proc. Roy. Soc.*, **B129,** 307 (1940).

[22] Hill, A. V., *Proc. Roy. Soc.*, **B136,** 243 (1949).

[23] Hill, A. V., *Ibid.*, **B136,** 195 (1949).

[24] Hill, A. V., *Ibid.*, **B126,** 136 (1938).

[25] Meyer, K. H., and Mark, H., "Der Aufbau der hochpolymeren organischen Naturstoffe," Akademische Verlagsgesellschaft M.B.H., Leipzig, 1930.

# THE STRUCTURE OF FIBROUS PROTEINS OF THE COLLAGEN-GELATIN GROUP

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated March 31, 1951

Collagen is a very interesting protein. It has well-defined mechanical properties (great strength, reversible extensibility through only a small range) that make it suited to the special purposes to which it is put in the animal body, as in tendon, bone, tusk, skin, the cornea of the eye, intestinal tissue, and probably rather extensively in reticular structures of cells. During the last thirty years, following the pioneer work of Herzog and Jancke,[1] a number of investigators have attempted to find the structure of collagen (and of gelatin, which gives similar x-ray photographs), but no one has previously proposed any precisely described configuration, nor has attempted to account for the positions and intensities of the x-ray diffraction maxima.

The diffraction pattern of collagen and gelatin is characterized by a meridional arc at 2.86 A. (Good reproductions of x-ray photographs have been published by Astbury.[2]) This arc remains essentially uninfluenced by a change in the source of material or its previous treatment; Bear[3] found that it varied only between the limits 2.82 A and 2.90 A for 26 samples, ranging from demineralized mammoth tusk to plain surgical gut (sheep intestinal submucosa). On the other hand, the principal equatorial reflection, which for thoroughly dried tendon[2] corresponds to the spacing 10.4 A, varies greatly in spacing with source and treatment of the material. At ordinary humidity it is about 11.5 A, and Bear reported the value 15.5 A for kangaroo tail tendon treated with water. It is evident that collagen consists of molecules (polypeptide chains) extending along the fiber axis, and rather loosely packed in parallel orientation. It will be pointed out below that the equatorial reflections correspond to a hexagonal packing of circular cylinders.

The 2.86-A fiber-axis spacing suggests that the amide groups of the polypeptide chain are in the cis configuration, as has been mentioned by Astbury.[2] (Astbury has suggested[2, 4] a structure for collagen which bears little resemblance to our structure.) If we accept the configuration of the amide group in a peptide as predicted from x-ray investigations of related simple substances, and described in our earlier paper[5] (with a single change —we have replaced the value 120° for the C′—N—C angle by the value 123°, which is suggested by general considerations as a reasonable value for the angle between a single bond and a bond with about 50 per cent

double-bond character), we predict that the length of an amide group with the cis configuration is 2.83 A, which is close to the observed fiber-axis spacing. Although the principal meridional arc at 5.1 A observed for proteins with the $\alpha$-keratin structure, which had always been accepted as representing a fiber-axis distance, has been found[6, 7] to be in fact a diagonal spacing, there seems to be little reason to doubt that the 2.86 A spacing of collagen represents a fiber-axis distance.

We found it impossible to formulate a satisfactory structure for collagen

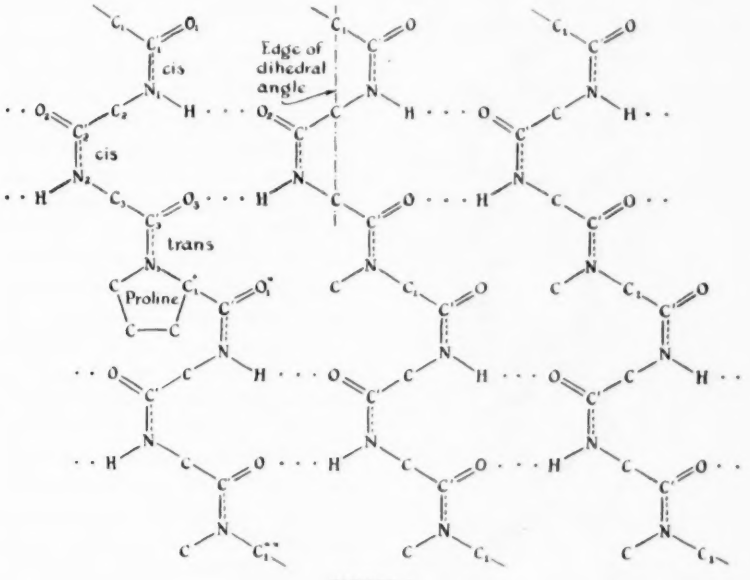



FIGURE 1

Diagrammatic representation of the configuration of polypeptide chains in the collagen-gelatin three-chain helix.

from cis amide groups alone. However, a satisfactory structure, described in detail in the following paragraphs, has been formulated with use of polypeptide chains in which there is an alternation of two cis groups and one trans group, as shown in figures 1 and 2. The angular orientation which the trans group is required to assume by its bonds with the contiguous cis groups is such as to cause its component of length along the fiber axis to be about 2.92 A, the average fiber-axis length per residue for the cis-cis-trans chain thus being 2.86 A, as observed for the collagen fibers.

Let us now consider the folding of the cis-cis-trans chain in such a way as to form satisfactory hydrogen bonds. The ease of lateral swelling of

collagen indicates that hydrogen bonds between molecules of the protein are not present, and hence that a structure involving intramolecular hydrogen bonds is to be sought.

The principal equatorial spacing for thoroughly dried collagen is 10.4 A. For cylindrical molecules in hexagonal packing this value for $d_{10\text{-}0}$ leads to 125 $A^2$ as the basal-plane area per molecule, and with 2.86 A as the fiber-axis length to 358 $A^3$ as the volume per unit, or 215 $cm^3$ per mole of units. If the density is taken as 1.35 g $cm^{-3}$ (the reported density of dry gelatin) the mass per unit is 291 g. The average residue weight for collagen and gelatin is found by analysis to be slightly less than 100; it is hence evident that there are about three residues in the unit of the molecule (2.86 A length). We accordingly reach the conclusion that the molecule of collagen and gelatin, essentially cylindrical in shape, is not a single polypeptide chain, but consists of three polypeptide chains. This is pleasing, inasmuch as there is no way in which a single polypeptide chain with a cis-cis-trans sequence of amide groups can be folded to give intramolecular hydrogen bonds and a fiber-axis residue length of 2.86 A. The intramolecular hydrogen bonds in the collagen-gelatin molecule must be lateral bonds between the three polypeptide chains that constitute the molecule.

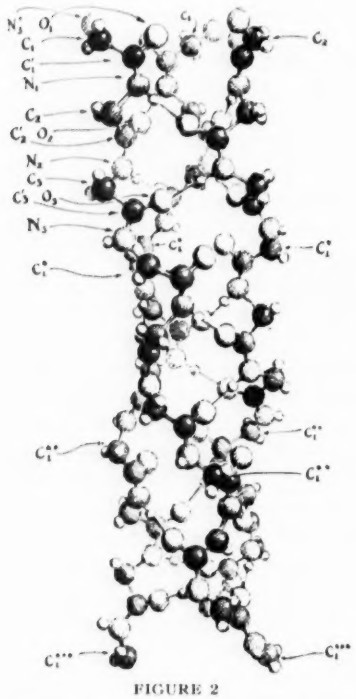



FIGURE 2

Drawing representing the proposed structure of the collagen-gelatin molecule.

A satisfactory structure can be built in which each of the three polypeptide chains is coiled into a helix, the coiling being achieved through the bending of the chain at the positions of the $\alpha$ carbon atoms, and the three helixes having a common axis. For the ideal configuration, in which the three $\alpha$ carbon atoms are similarly oriented directly above and below one another along a line parallel to the fiber axis, the dihedral angle at these carbon atoms is 97°. Some distortion of the structure is required in order that satisfactory hydrogen bonds be formed. This distortion consists in a

rotation around the two single bonds formed by the $\alpha$ carbon atoms, so as to draw the $\alpha$ carbon atom $C_2$ closer to the axis of the helix, by 0.34 A, than the $\alpha$ carbon atoms $C_1$ and $C_3$, as shown in figure 2. The hydrogen bonds are introduced in such a way that the three chains of the molecule are related to one another by a three-fold axis of symmetry. The $\alpha$ carbon atoms $C_2$ of the three chains lie at the corners of an equilateral triangle of edge 5.60 A, and the atoms $C_1$ and $C_3$ at the corners of an equilateral triangle with edge 6.20 A. The two cis amide residues are rotated by about 9° out of the orientation parallel with the fiber axis. The trans residue is rotated through 30° about its $C_3 - C_1^*$ axis. The trace of this axis on the basal plane is 2.44 A. The relation of this trace to the 6.20-A triangle of the $C_3$ and $C_1$ atoms is such that the angle of rotation about the fiber axis that converts one three-residue element of a polypeptide chain into the element following it in the chain is 40°. We have found that this angle of rotation can hardly be varied by more than 3° without introducing unsatisfactory structural features. The helix formed by a single chain of the collagen molecule is thus found to have very nearly a 9-fold screw axis of symmetry.

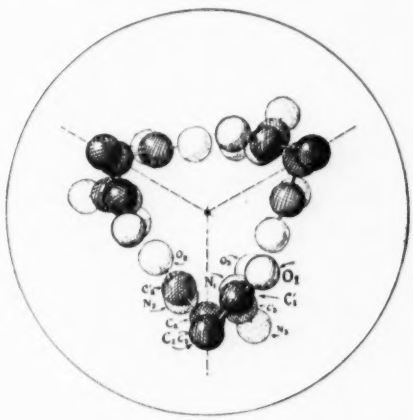



FIGURE 3

Plan of the three-chain configuration proposed for collagen and gelatin.

There is, of course, no reason for an isolated gelatin molecule to have exactly the angle 40° for the rotation of its rotatory translation, or to have exactly a 9-fold screw axis. In an aggregate of molecules in hexagonal packing, however, the influence of adjacent molecules on a given molecule might well be such as to introduce a small torque that would cause it to assume exactly the value 40° for this angle, and thus to assume a true 9-fold screw axis, in addition to the 3-fold symmetry axis that converts one chain into another, as shown in basal-plane projection in figure 3. We have found this phenomenon to occur in the hexagonal crystals of poly-$\gamma$-methyl-L-glutamate,[6] in which the $\alpha$ helix, normally with approximately 3.69 residues per turn, is constrained to the value $18/5 = 3.60$ in order to achieve a 6-fold screw axis. In the following discussion we describe a molecule with a 9-fold screw axis.

The translation of the operation that converts one structural element

into the following one in the chain is given for our model as 8.58 A, which corresponds to 2.86 A per residue, in exact agreement with the x-ray value.

The coordinates of atoms in a structural unit of the molecule are given in Table 1. The coordinates $x$, $y$, and $z$ refer to cartesian axes centered at the $\alpha$ carbon atom $C_2$, and the coordinates $\rho$, $\theta$, and $z$ are cylindrical coordinates referred to the axis of the helix. These atomic positions correspond closely to the assumed bond angles, bond distances, and planarity of the

TABLE 1

ATOMIC PARAMETERS FOR THE COLLAGEN THREE-CHAIN HELIX

$x$, $y$, $z$, AND $\rho$ IN A

| ATOM | $x$ | $y$ | $z$ | $\rho$ | $\phi$ |
|---|---|---|---|---|---|
| $C_1$ | −0.25 | −0.25 | 2.83 | 3.59 | 0.0° |
| $C_1'$ | 1.16 | −0.09 | 2.19 | 2.63 | 19.6° |
| $O_1$ | 2.18 | −0.09 | 2.90 | 2.38 | 42.3° |
| $N_1$ | 1.17 | 0.00 | 0.88 | 2.55 | 18.9° |
| $C_2$ | 0.00 | 0.00 | 0.00 | 3.25 | 0.0° |
| $C_2'$ | 0.05 | 1.24 | −0.90 | 2.47 | −19.9° |
| $O_2$ | 0.22 | 2.38 | −0.41 | 2.07 | −47.5° |
| $N_2$ | −0.04 | 1.06 | −2.21 | 2.63 | −17.2° |
| $C_3$ | −0.25 | −0.25 | −2.83 | 3.59 | 0.0° |
| $C_3'$ | 1.01 | −0.37 | −3.73 | 2.94 | 19.3° |
| $O_3$ | 2.03 | 0.21 | −3.41 | 2.10 | 37.9° |
| $N_3$ | 0.89 | −1.09 | −4.80 | 3.66 | 22.5° |
| $C_1$* | 1.98 | −1.32 | −5.75 | 3.59 | 40.0° |
| Axis | 2.29 | 2.29 | ... | 0.00 | ... |

TABLE 2

COMPARISON OF EQUATORIAL FEATURES OF COLLAGEN X-RAY PHOTOGRAPHS AND CALCULATED MAXIMA AND MINIMA OF THE FORM FACTOR

| CALCULATED SPACINGS | | OBSERVED SPACINGS | |
|---|---|---|---|
| 7.14 A, | minimum intensity | 7.8 A, | center of light band |
| 4.72 | maximum | 4.37 | center of diffuse dark band |
| 2.89 | minimum | 2.74 | center of light band |
| 2.25 | maximum | 2.18 | center of dark band |
| 1.64 | minimum | | |

amide groups. Because of the complexity of the structure, we are not sure that a considerable displacement of the chains from the proposed positions might not exist, in such a way as to retain these structural features.

The nature of the inter-chain interactions is shown in figures 1 and 3. Two hydrogen bonds are formed per unit. The $N_1$—H···$O_2$ bond has the length 2.63 A. This is slightly shorter (by about 0.06 A) than any N—H···O hydrogen bond so far reported, but it seems to us to be possible that the bond in this molecule is indeed this short. The $N_2$—H···$O_3$ hydrogen bond has the length 2.83 A. If our model were to be changed by a small addi-

tional rotation of the trans amide group the distance would be significantly shortened.

There is an interesting correlation between the chemical composition of the collagen-gelatin proteins and the existence in the structure of only two, rather than three, hydrogen bonds per element. It is found by chemical analysis that about one-third of the amino acids obtained by hydrolysis of collagen and gelatin are proline or hydroxyproline. In residues of these amino acids the nitrogen atom does not have an attached hydrogen atom and so does not enter into hydrogen-bond formation. The proposed structure may explain this aspect of the chemical composition of these proteins. It seems likely that if long polypeptide chains were to be synthesized with every third residue a proline or hydroxyproline residue and every third residue a glycine residue they would spontaneously aggregate into complexes of three, with the collagen structure, the other principal known structures for fibrous proteins, the $\alpha$-helix structure, pleated-sheet structure, and $\gamma$-helix structure, being rendered unstable by the steric interference of the 5-membered rings that are held in a fixed orientation.

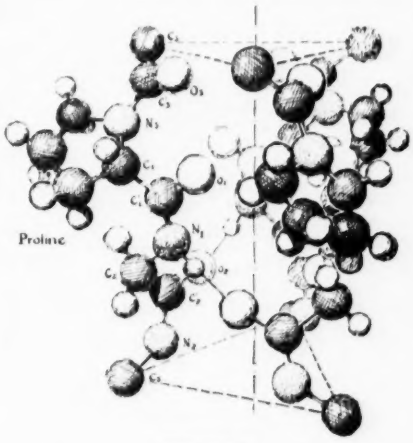



FIGURE 4

Drawing of a portion of the proposed structure of the collagen-gelatin molecule, showing the positions of the proline residues.

It is of interest that the sense of the collagen helix is uniquely related to the configuration of the amino acid residues. It is possible for only one of the two alternative positions of a $\beta$ carbon atom on the $\alpha$ carbon atom $C_1$ of the proline residue to be connected with the nitrogen atom to form a 5-membered ring, for a given sense of the helix. The acceptable configuration is shown in figure 4.

The structure of the collagen-gelatin molecule can be represented diagrammatically as shown in figure 5. The three helical polypeptide chains are shown projected onto the surface of a cylinder. It is interesting to note that the structure of the molecule provides an immediate explanation of the principal mechanical property of collagen, its extensibility over only a limited range. The effective fiber-axis length of the trans residue in the molecule as shown is 2.93 A. If, on the application of force, this residue were to be twisted into a parallel orientation, its effective length would be

3.83 A, the $C_3$—$C_1$* distance. This maximum increase in length corresponds to a 10 per cent extension for the molecule. It is likely, however,

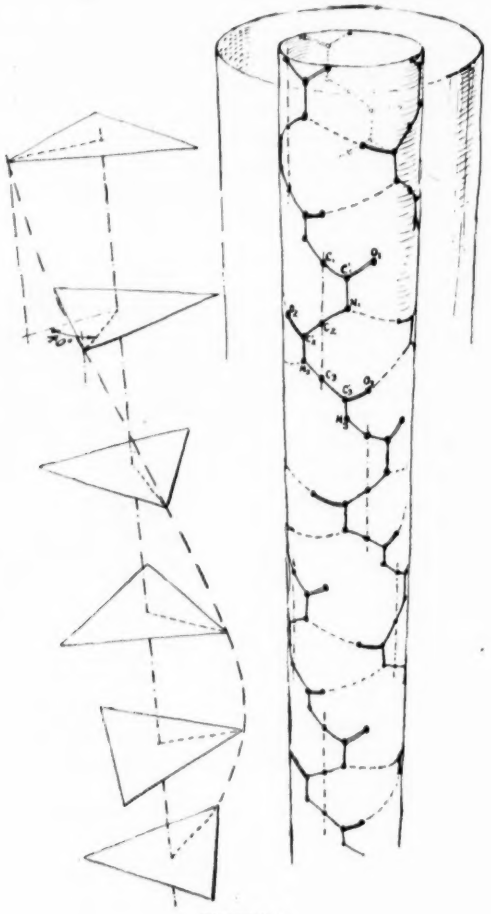


FIGURE 5

Diagrammatic representation of the collagen three-chain helix.

that the bond angles and planarity of amide groups would prevent complete parallel orientation, and that only a somewhat smaller extension could be achieved.

The way in which the molecules are packed together in a fiber of tendon or connective tissue is shown in figure 6. The cylindrical molecules are arranged in hexagonal packing. As discussed above, they are all helical, with the same sense, either right-handed or left-handed. It cannot be predicted whether in tendon or other collagenous material the molecules would all be oriented similarly, or whether some would be oriented in one direction and others in the opposite direction. It seems likely that the distribution of side chains reflects the direction of the polypeptide chains in the molecules, in such a way that the packing would be better in case that the molecules are all similarly oriented than if they were to alternate in

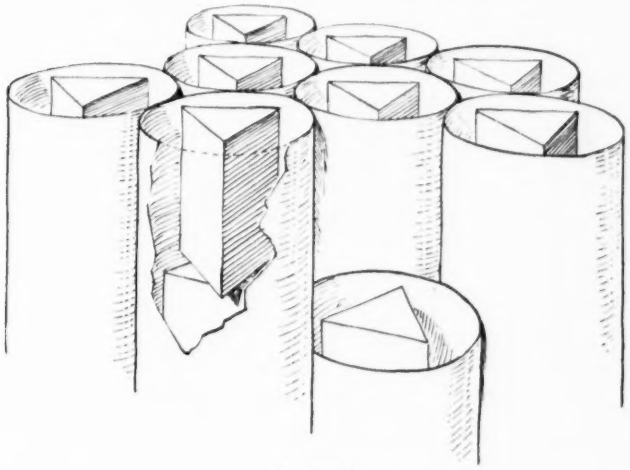


FIGURE 6

Diagrammatic representation of an aggregate of collagen molecules in a collagen fiber.

orientation. A preliminary study of the packing of side chains indicates that the glycine residues lie immediately below the proline and hydroxyproline residues (in the molecule oriented as shown in figure 2); that is, that the carbon atom $C_2$ is a methylene carbon atom of a glycine residue. It has been found by analysis that about one-third of the residues are glycine residues.

The proposed structure of the collagen-gelatin molecule accounts in a striking way for the principal features of the x-ray diffraction pattern of collagen and gelatin. A fiber consisting of these cylindrical molecules in hexagonal packing, with $a_0 = 12.50$ A (at normal humidity), would be predicted to produce equatorial reflections $\{10{\cdot}0\}$, with $d = 10.83$ A, $\{11{\cdot}0\}$,

with $d = 6.25$ A, $\{20\cdot0\}$, with $d = 5.42$ A, and $\{22\cdot0\}$, with $d = 3.13$ A. The lines reported by Corey and Wyckoff,[8] Astbury,[2] and other investigators are a strong reflection at 10.9 A and a medium reflection at 5.42 A. We may ask why these reflections appear, and $\{20\cdot0\}$ and $\{22\cdot0\}$ do not. The answer is given by the calculation of the form factor for the collagen molecule, with the parameters of table 1, and the assumption of cylindrical symmetry. The equation $F = \sum_i f_i J_0 (4\pi\rho_i \sin\theta/\lambda)$ leads to the function $F$ shown in figure 7. This function has nodes and maxima as given in table 2, where comparison is made with the corresponding features as measured by us on the photographs of raw kangaroo tendon made by Corey and

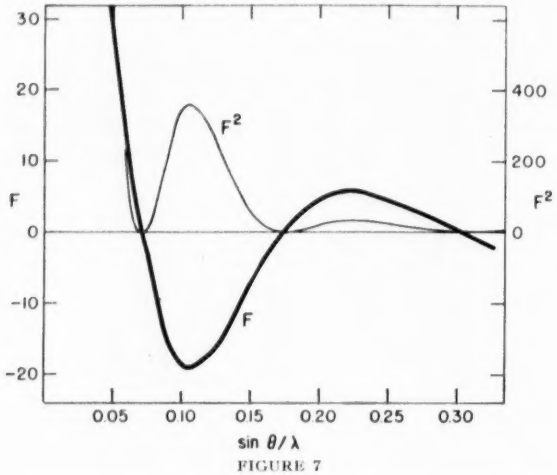



FIGURE 7

Calculated form factor and square of form factor for equatorial reflections of the collagen three-chain helix, calculated for cylindrical symmetry.

Wyckoff.[8] We see that it is predicted that a minimum in intensities of reflections would occur at about $d = 7.14$ A and another minimum at $d = 2.89$ A; these values are close to those for $\{20\cdot0\}$ and $\{22\cdot0\}$, 6.25 A and 3.13 A, respectively. Moreover, there is observable on the photographs, and on published photographs of collagen, such as those of Corey and Wyckoff and of Astbury, a general light band, at about the azimuthal angle predicted for the minimum at 7.14 A equatorial spacing, followed by a region of general darkening, for which Astbury has located the center at 4.4 A, and which we have also measured at 4.37 A. This band is followed by a lighter band, and then a faint darker ring, the center of which we measure at 2.18 A, in excellent agreement with the predicted maximum of inten-

sity, 2.25 A. It seems likely that the presence of the 3-chain collagen helix can be recognized in a fibrous material more easily from this general blackening, arising from disordered molecules, than from the crystallographic diffraction maxima.

The predicted unit of structure of crystalline collagen, except for perturbations due to the distribution of side chains, is a hexagonal unit with $a_0$ = 12.5 A (for collagen at normal humidity) and $c_0$ = 25.74 A, the fiber-axis distance for nine residues. We have noted that the wide-angle pattern of collagen can be completely or nearly completely indexed in terms of this unit. Except for the side chains, the suggested hexagonal structure involves only one undetermined parameter, the azimuthal angle fixing the orientation of the molecules relative to the crystal axes. The meridional small-angle reflections reported by Corey and Wyckoff,[8] Bear,[3] and other workers are also seen to be related to the unit. These reflections seem to represent orders from the large spacing of about 640 A (which appears not only in x-ray photographs but also in electron micrographs[9]), apparently enhanced when they bear a nearly rational relation to $c_0$.

(Added May 5, 1951.) An interesting paper on infrared spectra and structure of fibrous proteins, by E. J. Ambrose and A. Elliott, *Proc. Roy. Soc.*, **A 206,** 206 (1951), has just appeared. The results obtained agree well with our proposed structures. In particular, Ambrose and Elliott conclude that in collagen "the N—H bond must be nearly if not exactly normal to the chain axis, and this applies also to the C=O bond," in excellent substantiation of the prediction on the basis of our three-chain helical structure.

This work was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The U. S. Public Health Service.

* Communication No. 1555.

[1] Herzog, R. O., and Jancke, W., *Ber.*, **53,** 2162 (1920).
[2] Astbury, W. T., *J. Int. Soc. Leather Trades' Chem.*, **24,** 69 (1940).
[3] Bear, R. S., *J. Am. Chem. Soc.*, **66,** 1297 (1944).
[4] Astbury, W. T., and Bell, F. O., *Nature*, **145,** 421 (1940).
[5] Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37,** 205 (1951).
[6] Pauling, L., and Corey, R. B., *Ibid.*, **37,** 241 (1951).
[7] Pauling, L., and Corey, R. B., *Ibid.*, **37,** 261 (1951).
[8] Corey, R. B., and Wyckoff, R. W. G., *J. Biol. Chem.*, **114,** 407 (1936).
[9] Schmitt, F. O., Hall, C. E., and Jakus, M. A., *J. Cell. Comp. Physiol.*, **20,** 11 (1942); *J. Am. Chem. Soc.*, **64,** 1234 (1942).

# THE POLYPEPTIDE-CHAIN CONFIGURATION IN HEMOGLOBIN AND OTHER GLOBULAR PROTEINS

By Linus Pauling and Robert B. Corey

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology, Pasadena, California

Communicated March 31, 1951

In the immediately preceding papers we have described several hydrogen-bonded planar-amide configurations of polypeptide chains, and have discussed the evidence bearing on the question of their presence in fibrous proteins. It seems worth while to consider the possibility that these configurations—the pleated sheet, the 3.7-residue $\alpha$ helix, the 5.1-residue $\gamma$ helix, and the three-chain collagen helix—are represented in molecules of the globular proteins.

It may first be noted that many globular proteins, such as ovalbumin, can on denaturation be converted into a form showing the $\beta$-keratin x-ray pattern.[1] The fiber-axis residue distance that is observed, about 3.3 A, is the same as for $\beta$ keratin, for which we have suggested the pleated-sheet configuration,[2] and it seems reasonable that the same structure should be represented by these denatured proteins. It is, of course, to be expected that a layer structure, such as the pleated sheet, would be assumed by a protein when pressed flat, and the extension of the chains in the pleated-sheet structure makes it reasonable that such a structure should also be assumed by a protein when drawn into a fiber.

The most significant published data bearing on the structure of globular proteins are those on horse carbonmonoxyhemoglobin that have been obtained through the well-planned and diligent efforts of Perutz and his co-workers.[3, 4] These data have been published mainly as a set of sections of a three-dimensional Patterson diagram. We have observed that the data provide some support for the idea that the 3.7-residue helix is a principal feature of the structure of this protein.

Perutz has pointed out that his data indicate that the hemoglobin molecule is about 57 A long, and between 34 A and 57 A in other dimensions, and that there are present rods extending in the 57-A direction, and packed in a pseudohexagonal array, with the centers of the rods about 10.5 A apart. He concluded that the rods probably have the same structure as the molecules in $\alpha$ keratin, for which we have recently suggested the 3.7-residue helical configuration.[5]

There are several facts that favor the view that the 3.7-residue helix is represented in hemoglobin. First, there is the similarity to $\alpha$ keratin, pointed out by Perutz, and the evidence supporting the 3.7-residue helical configuration for the fibrous proteins with the $\alpha$-keratin structure.[5] Closely related is the fact that from the density and the average residue weight for

hemoglobin one would predict that molecules with this helical configuration would be spaced about 11 A apart (from center to center), in agreement with Perutz's conclusion that the rods in hemoglobin are about 10.5 A apart. (A calculation of this sort at once eliminates the 5.1-residue helix, for which the predicted average spacing of the rods is 14 A.)

Another bit of supporting evidence is provided by the integrated vector density in a strip of the $xz$ Patterson section through the origin of the 3-dimensional diagram and in the direction of the axes of the rods. Bragg, Kendrew, and Perutz[6] have reproduced this quantity, plotted as a function of the distance from the origin, in connection with their painstaking analysis

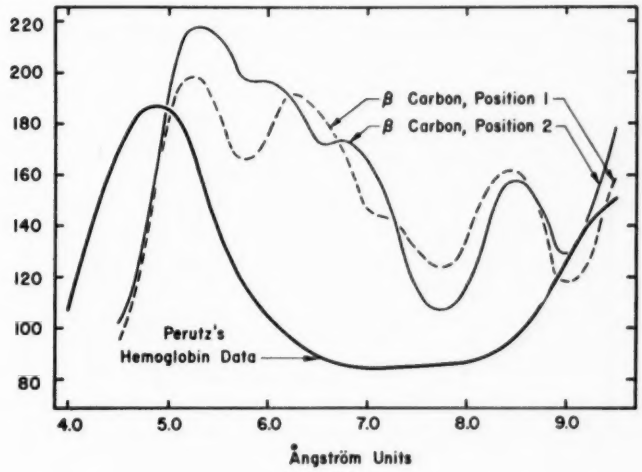



FIGURE 1

Comparison of the radial distribution function calculated for the 5.1-residue helical configuration, with inclusion of a $\beta$ carbon atom per residue, and the experimental radial distribution function for carbonmonoxyhemoglobin, as calculated from the three-dimensional Patterson function given by Perutz.

of the data for hemoglobin and also for myoglobin[7] and discussion of the correlation of the data with alternative polypeptide configurations. The function has peaks at about 5 A, 11.5 A, 16.5 A, 21.5 A, 27 A, 32 A, etc. We have evaluated a corresponding function for the 3.7-residue helix by including interatomic vectors deviating by not more than 2 A from the direction of the helical axis, and weighting the vectors proportionately to the product of the atomic numbers of the two atoms. The function obtained in this way for an 18-residue 5-turn helix with fiber-axis residue

length 1.53 A has maxima at 5.1 A, 10.6 A, 16.7 A, 21.4 A, 27.5 A, 32.6 A, etc., in excellent agreement with the experimental points.

Another test of the proposed configuration can be made by comparison of the calculated and observed radial distribution functions. Perutz pointed out that the Patterson diagram shows a strong shell at about 5 A from the origin. We have obtained a radial distribution function corresponding to his data for hemoglobin by numerical integration over the contoured Patterson sections published in his paper; this function is shown in figures 1 and 2. It is seen that it has a maximum at about 4.8 A. The calculated radial distribution functions for the 5.1-residue helix are also shown in figure 1. The two curves represent respectively the function for

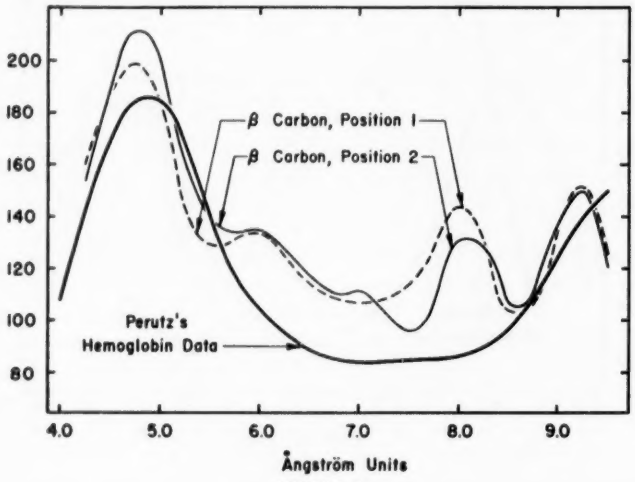



FIGURE 2

Comparison of the radial distribution function calculated for the 3.7-residue helical configuration of the polypeptide chain, with inclusion of a $\beta$ carbon atom for each residue, and the experimental radial distribution function for hemoglobin.

the four main-chain atoms C, C′, O, and N and a $\beta$ carbon atom in one of the two alternative positions, and the function for the four main-chain atoms and a $\beta$ carbon atom in the other position. It is seen that there is no agreement with the hemoglobin curve. The same two calculated radial distribution functions for the 3.7-residue helix are given in figure 2. We think that the rough agreement with the hemoglobin curve is to be considered as significant; it is to be remembered that even with inclusion of the $\beta$ carbon atom only about 60 per cent of the heavy atoms in the molecule have been taken into consideration in the calculation. The neglected

side-chain atoms are, of course, far more randomly arranged than the main-chain atoms of the helix, and would for this reason tend to distribute their vectors rather uniformly, and thus not to mask the characteristic features of the function due to the main-chain and $\beta$ carbon atoms.

The comparison of radial distribution functions may thus be construed as giving additional evidence in favor of the suggestion that the rods that Perutz has reported to be present in the hemoglobin molecule have the 3.7-residue helical configuration.

We think that it is not unlikely that this polypeptide configuration is represented in other globular proteins also. In particular, its presence in myoglobin, which is closely related to hemoglobin, would not be surprising; however, it must be pointed out that the Patterson projection for myoglobin on a plane perpendicular to the axis of the rods, given by Bragg, Kendrew, and Perutz,[6] seems hardly to be compatible with this structure. It is possible, of course, that side-chain atoms happen to cooperate effectively in changing the aspect of this projection, or that the axes of the rods do not lie exactly along the direction of projection. The evidence favoring the 3.7-residue helix for myoglobin is contained in Kendrew's description of the myoglobin molecule, as deduced from his data, as consisting of a layer of four rods about 9.5 A apart and with vector maxima spaced 5 A apart in the direction of the axes of the rods. The layers themselves are about 15 A apart, which suggests that if the structure does involve the 3.7-residue helix the side chains are distributed as in crystalline muscle,[5] in which the molecules have an effectively elliptical cross-section, with major and minor diameters 13.1 A and 9.8 A, respectively.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The United States Public Health Service. We acknowledge with gratitude the assistance and encouragement of our colleagues in The Gates and Crellin Laboratories of Chemistry throughout the period during which the studies reported in this series of papers and also the investigations on which this work is based were made. We are especially grateful to Professor Verner Schomaker, who has helped by giving us the benefit of both his deep understanding of structural chemistry and his profound critical insight.

* Contribution No. 1556.

[1] Astbury, W. T., and Lomax, R., *J. Chem. Soc.*, **1935,** 846; Astbury, W. T., Dickinson, S., and Bailey, K., *Biochem. J.*, **29,** 2351 (1935).

[2] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 251 (1951).

[3] Boyes-Watson, J., Davidson, E., and Perutz, M. F., *Proc. Roy. Soc.*, **A191,** 83 (1947).

[4] Perutz, M. F., *Ibid.*, **A195,** 474 (1949).

[5] Pauling, L., and Corey, R. B., these PROCEEDINGS, **37,** 261 (1951).

[6] Bragg, W. L., Kendrew, J. C., and Perutz, M. F., *Proc. Roy. Soc.*, **A203,** 321 (1950).

[7] Kendrew, J. C., *Ibid.*, **A201,** 62 (1950).

# THE INTERACTION OF MUCOPROTEIN WITH SOLUBLE COLLAGEN; AN ELECTRON MICROSCOPE STUDY*

BY JOHN H. HIGHBERGER, JEROME GROSS AND FRANCIS O. SCHMITT

RESEARCH DIVISION, UNITED SHOE MACHINERY CORPORATION, BEVERLY, MASSACHUSETTS; MEDICAL CLINIC OF THE MASSACHUSETTS GENERAL HOSPITAL;† AND BIOLOGY DEPARTMENT, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated March 25, 1951

The collagen of certain forms of connective tissue, such as rat tail tendon and the fish swin bladder (ichthyocol), dissolve in dilute acid to yield a clear, relatively viscous solution. When NaCl is added to such a solution to a concentration of 0.2–1.0%, or if the solution be neutralized, a fibrous precipitate of collagen is produced.[1–3] Electron microscope studies have demonstrated that the reconstituted fibrils show the axial period and intraperiod fine structure typical of native collagen fibrils although the acid filtrate contains only very thin filaments.[4, 5] The process by which the thin filaments in the acid filtrate aggregate laterally to produce the typical collagen structure is of interest not only from the physical chemical point of view but also because a better understanding of the phenomenon may provide clues as to the mechanism of fibrogenesis *in vivo*. Investigations of the process of fibril reconstitution from acid filtrates of collagen by the addition of salt have been made in these laboratories[6] and will be reported in detail elsewhere. For the present it may be noted that the type of fibril structure observed in the electron microscope (axial repeating patterns of about 650 A, 220 A, or no apparent pattern) depends upon the concentrations of salt and collagen. The experiments described in this paper suggest that other factors may also be of importance in the process of reconstitution.

Most forms of dense connective tissue, such as the corium of the skin, swell to varying extents in dilute acid but do not appreciably dissolve. The important discovery was made recently by a group of Russian workers[7–9] that if mammalian skin corium is extracted with organic acid buffers (chiefly citrate, pH 3.0 – 4.5, $\Gamma/2 = 0.2$) the extract, after dialysis against water, yields a "crystalline" fibrous material. This material they have called "procollagen," believing that it may be biochemical precursor of collagen. From physical chemical studies Bresler, *et al.*,[9] deduced that procollagen is a cylindrical macromolecule having a diameter of about 17 A, a length of 380 A and a weight of about 70,000. It was suggested that the macromolecule is composed of a coiled polypeptide chain about 2400 A long and that fiber formation is due to linear and lateral aggregation of procollagen macromolecules.

Electron microscope examination of "procollagen" prepared according

to the directions of Orekhovich, *et al.*,[8] revealed two types of fibrils: cross-striated collagen-type fibrils having an axial period averaging about 650 A, and a new type of fibril having an axial period ranging from 2000 to 3000 A.[10] Until this new type of fiber is characterized chemically it will be referred to as "long-spacing" or, briefly, LS fibril.

LS fibrils were obtained from extracts of rat, calf and steer skins, rat tail tendon and fish swim bladder. Considerable fine structure, consisting of characteristically positioned bands, may be observed within the main period (figure 3). An analysis of this intraperiod structure will be presented subsequently. The period is fairly uniform within individual fibrils but varies considerably in different fibrils in the same preparation. The distribution curves of preparations from various tissues differ somewhat

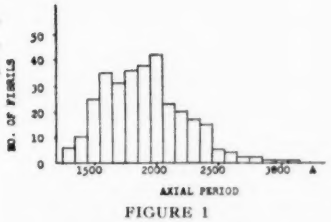



FIGURE 1

Highberger, J. H., Gross, J. and Schmitt, F. O.

Distribution of main axial periods of LS fibrils from dialyzed citrate extract of skin from 3 months old rat (314 fibrils were measured).

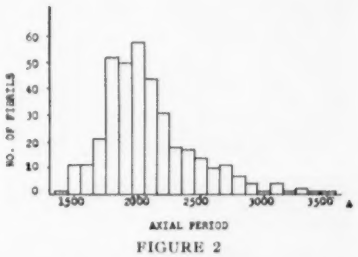


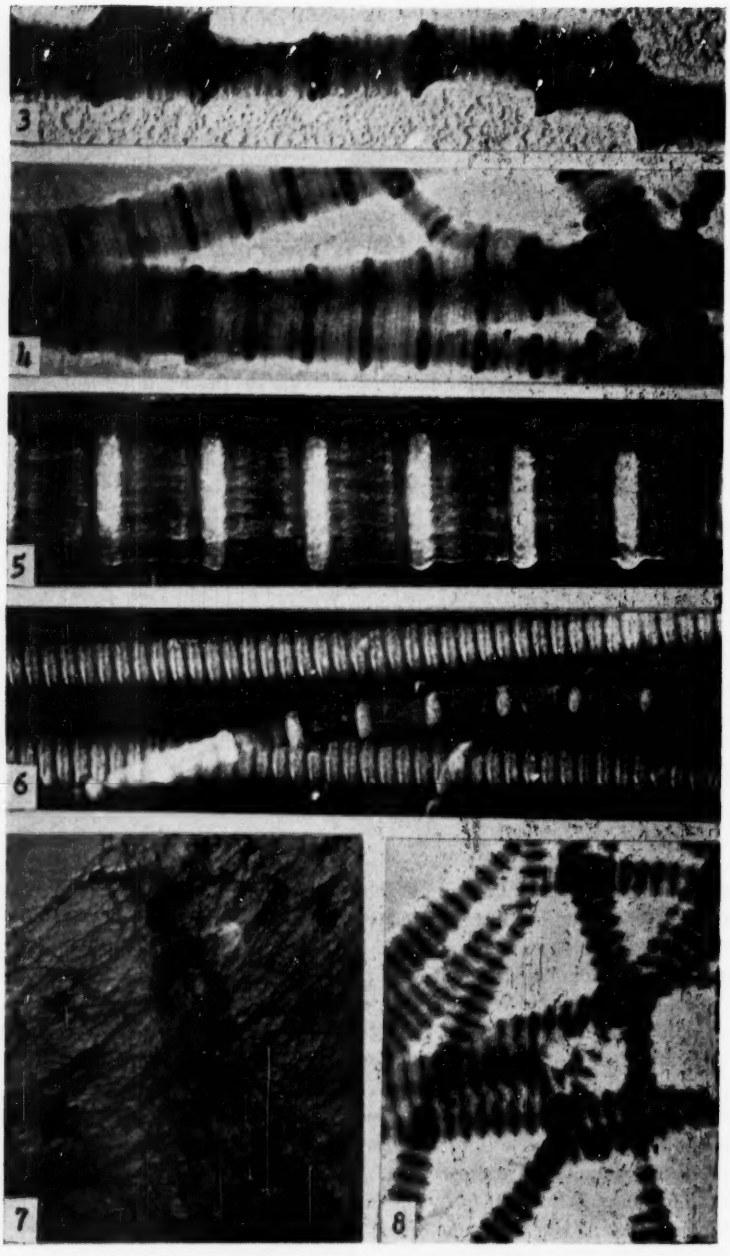

FIGURE 2

Distribution of main axial periods of LS fibrils from dialyzed mixture of plasma mucoprotein and acetic acid solution of ichthyocol (370 fibrils were measured).

among themselves. In figure 1 is shown the distribution curve of axial long spacings observed in material prepared from rat skin.

No LS fibrils have been observed in the numerous preparations of fragmented native collagen fibers which have been studied in this laboratory over a period of some years. Nor have they been observed thus far in thin sections of connective tissue though it must be admitted that relatively little study of thin sections has been made as yet upon skin and tendon (which are good sources of LS fibrils). Though it is possible that the fragmentation and sectioning techniques may destroy LS fibrils, it seems more probable that this new fibrous form does not pre-exist in the tissue but is produced artificially from precursors present in connective tissue extracts.

Collagen may be purified by the method of Bergmann and Stein,[11] which includes extensive extraction with 10% NaCl followed by extraction with dibasic phosphate and, finally, with ether. Such purified collagen yields




(See opposite page for caption).

very few if any LS fibrils when tested in the usual manner. It would appear that the purifying extractions remove a precursor or precursors of LS fibrils. Of the extractives used only the phosphate is involved in removal of precursor. Phosphate extracts of calf skin, rat skin and rat tail tendon, when dialyzed against water yield a flocculent precipitate, amorphous in the electron microscope. Extraction of this precipitate with citrate buffer (pH 3.8, $\Gamma/2 = 0.2$) and dialysis of the clear solution against water yields typical LS fibrils. Efforts were made, therefore, to determine the nature of the material which, with collagen, is extracted by phosphate and transformed into LS fibrils. Since mucoprotein (MP) has long been considered to be associated with collagen in the genesis of connective tissue and since phosphate is an excellent solvent for MP, the working hypothesis was tentatively adopted that MP may be one of the precursors of LS fibrils.

To test this hypothesis an MP preparation was added to a filtered acid solution of collagen obtained from tissue which had been sufficiently extracted with phosphate to remove precursor of LS fibrils. In relation to the bearing of the hypothesis on the genesis of connective tissue it might be most appropriate to use tissue MP for these tests. One experiment was in fact made with an MP preparation obtained from calf skin by a modification of the method of Weimer, *et al.*,[12] with results essentially similar to those described below. However, for most of the experiments plasma MP was used.

Bovine mucoprotein was isolated from beef plasma by the procedure described by Weimer, *et al.*[12] When added to an acetic acid filtrate of ichthyocol or rat tail tendon[13] in a ratio of 1 part of MP to about 10 parts of colla-

PLATE 1

Figure 3. LS fibrils from dialyzed citrate extract of skin from 3 day old rat. Stained with 1% phosphotungstic acid (PTA) and lightly shadowed with chromium. Magnification 74,000 ×.

Figure 4. LS fibrils precipitated by dialysis from a mixture of plasma mucoprotein (Schmid preparation) and acetic acid solution of ichthyocol. Stained with PTA. Magnification 45,000 ×.

Figure 5. LS fibril precipitated by dialysis from a mixture of plasma mucoprotein (Schmid preparation, total concentration 0.03%) and acetic acid solution of ichthyocol. Note longitudinal fibrillation. Shadowed with chromium. Magnification 68,000 ×.

Figure 6. Collagen-type fibrils (and one LS fibril) precipitated by dialysis from a mixture of plasma mucoprotein (total concentration 0.01%) and acetic acid solution of ichthyocol. Shadowed with chromium. Magnification 44,000 ×.

Figure 7. Fibrous gel formed by dialysis of acetic acid solution of ichthyocol. Note absence of striated fibrils. Stained with PTA. Magnification 13,000 ×.

Figure 8. LS fibrils precipitated by dialysis of a mixture of plasma mucoprotein (total concentration 0.2%) and acetic acid solution of ichthyocol (same solution which, when dialyzed without addition of mucoprotein yielded gel shown in figure 7). Stained with PTA. Magnification 13,000 ×.

gen a heavy fibrous precipitate was produced only after dialysis against water. This precipitate consisted almost entirely of typical LS fibrils (figure 8). The intraperiod fine structure (figure 4) and the distribution curve of the axial periods (figure 2) are closely similar to those typical of LS fibrils obtained from citrate extracts of the original tissue. Since our preparations of MP had not been highly purified, the experiment was repeated with an acid glycoprotein from human plasma, described by Schmid,[14] which was highly purified and proven to be monodisperse.[15] This material, the composition of which agrees fairly closely with that described by Weimer, *et al.*, produced LS fibrils identical with those produced by our own preparations of MP under similar experimental conditions. It seems safe to conclude, therefore, that it is in fact the MP, or some component of MP, which combines with acid-soluble collagen to produce LS fibrils. It should be noted that this reaction of MP with collagen appears to occur only when the collagen is highly solvated or dissolved in acid. Addition of MP to intact, native collagen fibrils does not produce LS fibrils.

When MP is added to acid filtrates of rat tail tendon or ichthyocol collagen in ratios of the order of 1 : 1000 or less, few if any LS fibrils are produced after dialysis against water. Rather, large numbers of fibrils having an axial period and intraperiod fine structure similar to collagen are produced. At intermediate ratios, mixtures of LS fibrils and the collagen type occur (figure 6). It is unlikely that this could be a simple reconstitution of collagen fibrils from the acid filtrate for this would require either the presence of salt or neutralization. Dialysis of collagen acid filtrates against water produces a gel consisting primarily of very fine filaments (figure 7). In figure 8 is shown a typical field of LS fibrils produced from the same preparation of ichthyocol but to which a relatively high concentration of MP had been added before dialysis.

These results suggest that MP, or a component of MP,[16] may be involved in the formation of collagen fibrils. When the MP content exceeds a critical value, LS rather than collagen fibrils are produced *in vitro*. At present, the mechanism of the reaction between MP and acid-soluble collagen is itself unknown. Detailed analytical investigations of this complex and, indeed, of reconstituted and native fibrils themselves, will be required before the above suggestion can be tested experimentally. Investigations along these lines are in progress.

* This investigation was supported in part by grants from the Eli Lilly Company to the Massachusetts Institute of Technology and by the United States Public Health Service to the Massachusetts General Hospital.

† Robert W. Lovett Memorial Foundation for the Study of Crippling Disease, Department of Medicine, Harvard Medical School.

[1] Nageotte, J., and Guyon, L., *Arch. Biol.*, **41**, 1–35 (1930).

[2] Huzella, T., *Z. Krist.*, **83**, 89–96 (1932).

[3] Faure-Fremiet, E., and Garrault, H., *Arch. Anat., mic.*, **33**, 81–94 (1937).

[4] Schmitt, F. O., Hall, C. E., and Jakus, M. A., *J. Cell. and Comp. Physiol.*, **20,** 11–33 (1942).

[5] Bahr, G., *Experimental Cell Research*, **1,** 603–606 (1950).

[6] Schmitt, F. O., and Gross, J., Unpublished.

[7] Tustanovsky, A. A., *Biokhimiya*, **12,** 285–290 (1947).

[8] Orekhovich, V. N., Tustanovsky, A. A., Orekhovich, K. D., and Plotnikova, N. E., *Biokhimiya*, **13,** 55–60 (1948).

[9] Bresler, C. E., Finogenov, P. A., and Frenkel, S. Y., *Reports Acad. Sci. U.S.S.R.*, Moscow, **72,** 555–558 (1950).

[10] Highberger, J. H., Gross, J., and Schmitt, F. O., *J. Am. Chem. Soc.*, **72,** 3321 (1950).

[11] Bergmann, M., and Stein, W. H., *J. Biol. Chem.*, **128,** 217–232 (1939).

[12] Weimer, H. E., Mehl, J. W., and Winzler, R. J., *J. Biol. Chem.*, **185,** 561–568 (1950).

[13] The acid filtrates of collagen, as prepared in these experiments, contain practically no undissolved collagen fibrils (which had passed through the filter). No fibrillar material was observed in solutions of MP alone, before or after dialysis.

[14] Schmid, K. J., *Am. Chem. Soc.*, **72,** 2816 (1950).

[15] We are greatly indebted to Dr. Karl Schmid, of the Harvard Medical School, for supplying this material.

[16] The acid polysaccharides, hyaluronic acid and heparin precipitate acid-soluble collagen immediately upon mixing but no LS fibrils have thus far been found in this fibrous precipitate. Of two samples of protein-free chondroitin sulfate, kindly furnished by Dr. Karl Meyer, of the College of Physicians and Surgeons, Columbia University, one produced LS fibrils while the other did not. Further information is required to clarify this matter. No evidence is as yet available concerning the possible role of the protein component of MP, which may be an $\alpha_1$ globulin. Experiments on this subject are in progress.

## *AN HYPOTHESIS OF PROTEIN SYNTHESIS POTENTIATED BY CITRIC ACID**

BY SIDNEY W. FOX

CHEMICAL LABORATORY, IOWA STATE COLLEGE, AMES

Communicated by Henry Gilman, March 27, 1951

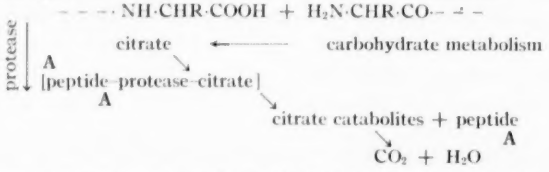
The need for understanding protein synthesis is sufficiently great to warrant the presentation of any hypothesis which is frankly recognized as such in the accurate meaning of this term, and which is not intrinsically untenable. The hypothesis presented here concerns the potentiation of peptide bond synthesis by citric acid metabolites. The essential suggestion arises from observations on the effect of increasing concentrations of citrate buffer on the papain-catalyzed synthesis of acylamino acid anilides.[1, 2]

The energy barrier in peptide bond synthesis has been evaluated as a figure within the range of 1400–3700 cal.[3] and this approximate range of figures has been used in a number of interpretations. Some of these statements are so limiting as to exclude reversal of proteolysis as a mechanism

of biological synthesis of peptide bonds (for a variety of inferences, cf. references 4–10). The values employed in such considerations have been calculated from heat capacity and heat of combustion values for solid amino acids and peptides. While it is theoretically sound, of course, to extend such values to aqueous systems, physiological fluids do not consist of pure water as solvent for such reactions, and if there are demonstrable effects of other solutes they must be considered.

The suggestion that citrate buffer may not be without biological significance in this connection, has been made;[1] this acid represents a natural metabolite. The effect of citrate concentration upon the speed of the forward synthetic reaction for anilides has been demonstrated to be a profound one,[1] and indicates that a similar effect should not be ignored in drawing conclusions on peptide bond formation, regardless of the type of intimate mechanism involved. While such experiments have been carried out with concentrations of substrate, enzyme, and buffer citrate that are many orders of magnitude greater than physiological values, the ratios of concentrations of enzyme to substrate to organic acid are not so unphysiological.

In considering more specifically how citrate might aid in this conversion, a reaction sequence such as the following deserves attention.




The forward reaction A would be joined to citrate as is the glassware counterpart; with citrate subject to catabolism directly from the complex, a favorable over-all equilibrium could result. The possibility of potentiation of peptide bond synthesis by coupling with energy-yielding reactions has been suggested before.[4] The present hypothesis provides a mechanism which does not invoke structural alteration of the intermediates. Such a sequence is thus compatible with the action of proteases, and with aqueous systems. It furthermore relies upon a type of coupling reaction for which a pertinent experimental model exists.

An hypothesis of this type is also compatible with some salient biological facts. Citric acid and related acids occur in cellular metabolism.[11] These organic acids occur as products of carbohydrate metabolism, which spills the great reservoir of energy in organisms.[12] It is of interest that the cytological location of the tricarboxylic acid cycle (e.g., mitochondria)[13] also represents sites of protein synthesis.[14, 15] It should perhaps be emphasized that more than one mechanism of protein synthesis may actually operate. Whether or not this is true, at least one acid known to be present

in cells has, in glassware, effects which cannot justifiably be ignored in the biological balance sheet, regardless of the validity of the specific hypothesis presented. Since a theoretical means for provision of the calculated energy requirement for the joining of carboxylic and aminoid peptide fragments exists, rejection of the proteolysis reversal concept of protein synthesis is not warranted.

The formulation of this hypothesis has been aided by critical discussions with Drs. George S. Hammond and Joseph F. Foster.

* Journal Paper No. J-1919 of the Iowa Agricultural Experiment Station, Project No. 1111. The work leading to this hypothesis has been supported in part by a grant from the National Cancer Institute of the National Institutes of Health, Public Health Service.

[1] Fox, S. W., and Pettinga, C. W., *Arch. Biochem.*, **25,** 13 (1950).

[2] Fox, S. W., and Wax, H., *J. Am. Chem. Soc.*, **72,** 5087 (1950).

[3] Huffman, H. M., *J. Phys. Chem.*, **46,** 885 (1942).

[4] Borsook, H., and Huffman, H. M., in Schmidt, C. L. A., *Chemistry of the Amino Acids and Proteins*, 2nd ed., Charles C Thomas, 1945, p. 866.

[5] Cohen, P. P., and McGilvery, R. W., *J. Biol. Chem.*, **171,** 131 (1947).

[6] Sprinson, D. B., *Isotopes in Biology and Medicine*, University of Wisconsin Press, 1948, p. 195.

[7] Northrop, J. H., Kunitz, M., and Herriott, R. M., *Crystalline Enzymes*, 2nd ed., Columbia University Press, 1948, p. 226.

[8] Sheehan, J. C., and Frank, V. S., *J. Am. Chem. Soc.*, **72,** 1312 (1950).

[9] Zamecnik, P. C., *Cancer Res.*, **10,** 659 (1950).

[10] Borsook, H., *Physiol. Revs.*, **30,** 206 (1950).

[11] Krebs, H. A., *Adv. Enzymol.*, **3,** 191 (1943); Wilcox, P. E., Heidelberger, C., and Potter, V. R., *J. Am. Chem. Soc.*, **72,** 5019 (1950).

[12] Muenscher, W. C., *Bot. Gaz.*, **75,** 249 (1923).

[13] Kennedy, E. P., and Lehninger, A. L., *J. Biol. Chem.*, **179,** 957 (1949).

[14] Brachet, J., *Chemical Embryology*, Interscience Publishers, 1950, p. 243.

[15] Haurowitz, F., *Chemistry and Biology of Proteins*, Academic Press, 1950, p. 340.

## *AN ADENINE REQUIREMENT IN A STRAIN OF DROSOPHILA**

BY TAYLOR HINTON, JOHN ELLIS,† AND D. T. NOYES

DEPARTMENT OF BIOLOGY, AMHERST COLLEGE, AMHERST, MASSACHUSETTS

Communicated by A. F. Blakeslee, March 1, 1951

With the improved growth now obtainable, the raising of *Drosophila melanogaster* upon a chemically defined medium under aseptic conditions,[1] renders possible more exacting studies of the nutritional requirements of genetically different strains. Thus a comparative study was made of a wild type (Oregon-R) and Inversion (2LR) 40d. The latter strain has been studied in detail in other respects[2, 3] and was found to possess a phenotype consisting of disarranged eye facets, mottled eye pigment, and

tumor-like black protrusions from the surface of the eye.[2] This phenotype was found to be directly correlated with the position of heterochromatin involved in a chromosomal inversion.[3] Furthermore, it was found that the abnormal phenotype could be caused to revert to the wild type, or normal, condition both when the position of the heterochromatin was changed by further chromosomal rearrangements[3] and by feeding the developing larvae an abundance of yeast at certain temperatures.[2] It appeared, therefore, that some factor in the food was needed by the inversion strain which was not required by the wild type in order to develop a normal eye. Since heterochromatin is also correlated with the development of the eye in the inversion strain, and since heterochromatin differs from euchromatin primarily in its nucleic acid synthesis,[4] it was suggested[2, 3] that the nucleic acid requirements of the two strains may differ. The present study was undertaken to test that hypothesis. Preliminary results were published earlier.[7]

*Technique.*—Eggs were collected and washed in 70% alcohol for 45 minutes. They were then inoculated, ten per tube, using bacteriological procedures, onto the surface of an agar slant of previously autoclaved medium. The base medium is given in table 4, and is a modification of the medium announced by Schultz, *et al.*[1] However, it should be noted that this medium is not in current use, a better medium having been devised.[5]

Growth was measured by an index derived by adding together the percentage of original larvae to pupate and the percentage of original larvae to become adults, and dividing that sum by the average time in days required for pupation. The larger the final figure so derived, the better the growth since the best growth would be in the case of a large percentage of pupation and eclosion in a short interval of time.

Each experimental set presented in the accompanying tables (except table 2) represents at least two independent experiments the results of which are added together.

*Nucleic Acid Omitted from the Diet.*—When the wild type strain was raised on medium containing no nucleic acid, the growth index was high (table 1), indicating, as has been reported by others,[1] that Drosophila does not require nucleic acid, normally being capable of synthesizing it. It may be noted, however, that growth is improved and the time required for pupation shortened upon the addition of nucleic acid to the medium.

On the other hand, the inversion strain proved to be incapable of completing development when nucleic acid was omitted from the diet (table 1). This clearly indicates that this strain, unlike the normal Drosophila, is incapable of synthesizing nucleic acid or some derivative of nucleic acid.

*The Nucleic Acid Derivative Required.*—Having established that the inversion strain failed to develop when nucleic acid was omitted from the diet, it remained to determine the nature of this requirement. The pre-

TABLE 1

| | | NO. LARVAE TO HATCH | PER CENT LARVAE TO PUPATE | AV. TIME TO PUPATION IN DAYS | PER CENT LARVAE TO BECOME ADULTS | GROWTH INDEX |
|---|---|---|---|---|---|---|
| Base medium plus nucleic acid (3 mg. per ml.) | wild type | 31 | 100 | 16.2 | 74.0 | 10.7 |
| | inversion | 11 | 90.9 | 19.3 | 72.7 | 8.5 |
| Base medium (no nucleic acid) | wild type | 92 | 82.6 | 20.1 | 63.0 | 7.2 |
| | inversion | 59 | 6.8 | 25.2 | 0.0 | 0.3 |

vious experiment does not show whether the whole nucleic acid molecule is required by the organism, or only one or more of its constituents. Therefore, larvae were raised upon media lacking nucleic acid, but containing, in each case, one purine or pyrimidine base, ribonucleoside, or ribonucleotide. All were tried with the exception of cytosine which was unavailable at that time. Wild type larvae were used in half of the tubes in each experiment to serve as controls, except in the case of uridine where all the tubes became contaminated and could not be counted. The results are presented in table 2.

Inspection of table 2 shows that the inversion strain gave a good growth index only when adenine (or its nucleoside or nucleotide) was present in

TABLE 2

| BASE MEDIUM PLUS | | NO. OF LARVAE | PER CENT LARVAE TO PUPATE | AV. TIME TO PUPATION DAYS | PER CENT LARVAE TO BECOME ADULTS | GROWTH INDEX |
|---|---|---|---|---|---|---|
| Guanine (0.11 mg. per ml.) | wild type (+) | 14 | 86 | 20.1 | 86 | 8.6 |
| | inversion | 28 | 32 | 22.3 | 21 | 1.3 |
| Guanosine (0.165 mg. per ml.) | + | 19 | 58 | 22.7 | 53 | 4.9 |
| | inv. | 23 | 0 | .. | 0 | 0.0 |
| Guanylic acid (0.25 mg. per ml.) | + | 59 | 85 | 20.8 | 56 | 6.8 |
| | inv. | 11 | 9 | 21.0 | 9 | 0.9 |
| Adenine (0.11 mg. per ml.) | + | 12 | 83 | 20.5 | 83 | 8.1 |
| | inv. | 20 | 60 | 17.1 | 55 | 6.7 |
| Adenosine (0.165 mg. per ml.) | + | 39 | 97 | 18.5 | 95 | 10.4 |
| | inv. | 20 | 60 | 15.7 | 60 | 7.6 |
| Adenylic acid (0.25 mg. per ml.) | + | 15 | 80 | 19.3 | 53 | 6.9 |
| | inv. | 14 | 71 | 10.8 | 57 | 11.8 |
| Uracil (0.11 mg. per ml.) | + | 46 | 65 | 21.2 | 43 | 5.1 |
| | inv. | 37 | 0 | .. | 0 | 0.0 |
| Uridine (0.165 mg. per ml.) | + | .. | .. | .. | .. | .. |
| | inv. | 7 | 0 | .. | 0 | 0.0 |
| Uridylic acid (0.25 mg. per ml.) | + | 39 | 74 | 23.1 | 67 | 6.1 |
| | inv. | 14 | 0 | .. | 0 | 0.0 |
| Cytidine (0.165 mg. per ml.) | + | 32 | 81 | 22.1 | 75 | 7.1 |
| | inv. | 11 | 0 | .. | 0 | 0.0 |
| Cytidylic acid (0.25 mg. per ml.) | + | 34 | 68 | 23.0 | 56 | 5.4 |
| | inv. | 56 | 0 | .. | 0 | 0.0 |

the medium. In all other cases the larvae failed to develop except with guanine (9 pupae and 6 adults) and with guanylic acid (1 pupa and 1 adult). Rather than suppose guanine is replacing adenine, it is more logical to assume that the guanine and guanylic acid used in these experiments have a slight contamination of adenine. This is especially logical since guanosine does not support growth, and it is known to be more pure while guanylic acid is known to contain, at least, pyrimidine contaminants.[6]

TABLE 3

| BASE MEDIUM PLUS BASES (0.11 MG. OF EACH, PER ML.) | | NO. OF LARVAE | PER CENT LARVAE TO PUPATE | AV. TIME TO PUPATION, IN DAYS | PER CENT LARVAE TO BECOME ADULTS | GROWTH INDEX |
|---|---|---|---|---|---|---|
| Guanine, Adenine, Uracil | wild type (+) | 29 | 86.2 | 17.7 | 55.2 | 8.0 |
| | inversion | 12 | 50.0 | 17.7 | 25.0 | 4.2 |
| Guanine, Uracil | + | 17 | 94.1 | 19.0 | 47.0 | 7.4 |
| | inv. | 10 | 0.0 | .. | 0.0 | 0.0 |
| PLUS RIBONUCLEOSIDES (0.165 MG. OF EACH, PER ML.) | | | | | | |
| Guanosine, Adenosine, Uridine, Cytidine | + | 48 | 85.4 | 15.3 | 54.2 | 9.1 |
| | inv. | 27 | 37.0 | 15.5 | 22.2 | 3.8 |
| Guanosine, Uridine, Cytidine | + | 34 | 64.7 | 22.5 | 20.6 | 3.8 |
| | inv. | 38 | 0.0 | .. | 0.0 | 0.0 |
| PLUS NUCLEOTIDES (0.25 MG. OF EACH, PER ML.) | | | | | | |
| Guanylic acid, Adenylic acid, Uridylic acid, Cytidylic acid | + | 33 | 90.9 | 15.9 | 42.4 | 8.4 |
| | inv. | 59 | 61.0 | 13.5 | 45.8 | 7.9 |
| Guanylic acid, Uridylic acid, Cytidylic acid | + | 38 | 65.8 | 20.6 | 26.3 | 4.5 |
| | inv. | 74 | 5.4 | 21.0 | 5.4 | 0.5 |

The wild type larvae gave fair growth indices (table 2) in all cases, although slight improvement was shown when adenine was present in some form and when guanine (with the supposed adenine contamination) was present. This leaves little doubt that the inversion strain differs in its nutritional pattern from the wild type in that it has lost the ability to synthesize adenine.

*Combinations of Purines and Pyrimidines.*—It remained to determine whether the adenine requirement of the inversion strain could be met by

any combination of purines and pyrimidines. Larvae were raised on media containing the purine and pyrimidine bases in all possible combinations; on media containing the purine and pyrimidine nucleosides in all possible combinations; and on media containing the purine and pyrimidine nucleotides in all possible combinations. The results of some of these are presented in table 3.

Table 3 can be summarized by pointing out that of all of the combinations of purines and pyrimidines tested, whether bases, nucleosides or nucleotides, none supported growth unless adenine (or adenosine, or adenylic acid) was in the combination. In no case where adenine (or

TABLE 4

BASE MEDIUM

| | MG. PER ML. | | MG. PER ML. |
|---|---|---|---|
| L-arginine | 0.794 | Cholesterol | 0.1 |
| L-cystine | 0.48 | Lecithin | 0.1 |
| L-glutamic acid | 4.418 | Sucrose | 5.0 |
| Glycine | 0.388 | $KH_2PO_4$ | 0.606 |
| L-histidine | 0.484 | $K_2HPO_4$ | 0.606 |
| DL-isoleucine | 1.258 | $CaCl_2$ | 0.0129 |
| L-leucine | 2.345 | NaCl | 0.0129 |
| L-lysine | 1.337 | $FeSO_4 \cdot 7H_2O$ | 0.0129 |
| DL-methionine | 0.339 | $MgSO_4 \cdot 7H_2O$ | 0.246 |
| L-phenylalanine | 1.008 | $MnSO_4 \cdot 4H_2O$ | 0.129 |
| DL-threonine | 0.756 | Agar | 15.0 |
| L-tryptophane | 0.349 | | |
| DL-valine | 1.355 | | |
| | GAMMA PER ML. | | GAMMA PER ML. |
| Biotin | 0.0094 | Pteroylglutamic acid | 6.0 |
| Choline chloride | 20.0 | Pyridoxine | 30.0 |
| *p*-Aminobenzoic acid | 2.0 | Riboflavin | 2.4 |
| Inositol | 42.0 | Thiamine | 1.5 |
| Niacinamide | 10.0 | Vitamin $B_{12}$ | 0.004 |
| Ca-pantothenate | 6.0 | Protogen | 2.5 units[8] |

adenosine, or adenylic acid) was omitted from the combination was a growth index higher than 0.6 obtained. In fact, in only two cases were any pupae obtained, and then only a total of 5. Since the results were consistent, data for the other combinations are not presented as they add nothing further to the conclusion.

It, therefore, is justifiable to conclude that the adenine requirement of the inversion strain cannot be satisfied by any combination of other purines and pyrimidines tested.

*Discussion.*—Since this is the first clear-cut case in animals of the inheritance of a basic biochemical difference involving a nutritional deficiency, it seemed fitting to publish the results at this time, even though the mode of inheritance cannot be conclusively stated.

As was pointed out at the beginning of the paper, it was the early studies on the behavior of the position effect known as In(2LR)40d which led to the initial attempts to search for differences in the nucleic acid requirement of this strain as compared to a wild type. That a difference was found does not prove that the difference (the adenine requirement) and the position effect are correlated, although the coincidence is highly suggestive. It can only be said, at this time, that in all crosses made the adenine requirement follows the chromosomal inversion in segregation, i.e., appears to be genetically linked to it. However, the techniques are complicated for distinguishing between the possibilities that the requirement is due to a gene associated with the inversion, and that it is an effect of the inversion itself. Nevertheless, a series of crosses to replace each chromosome is in progress, as well as a study of the requirements of reversions and partial reversions of the chromosomal inversion,[3] in an attempt to decide between these possibilities.

If the adenine requirement is due to the rearrangement of heterochromatin as a result of the inversion, it should then, in turn, be correlated with the phenotype manifested by the inversion. In the present work, phenotypic differences have been noted when the nucleic acid constituents of the medium were altered qualitatively or quantitatively. However, the natural variability of expression of this phenotype is great under most circumstances, and a true evaluation of the expression can be gained only by a statistical measurement of the mean degree of expression.[2] The present data are not extensive enough to make such a measurement valid. Also, the higher concentrations of nucleic acid in the diet become increasingly toxic thereby introducing the complicating factor of retarded growth. Any evaluation of the normal or abnormal development of the eye would have to correct for this factor. Experiments are in progress in an effort to clarify the nature of the correlation between adenine in the diet and the expression of the phenotype.

It seems certain that an adenine requirement is not a characteristic of all position effects. We have tested several different types of position effects without discovering another requirer. On the other hand, another adenine requirer has turned up in a strain which is thought to be cytologically normal.

When the strain requiring adenine is crossed with some other strains, the requirement usually behaves in the offspring as a dominant effect. However, when crossed to the Oregon-R wild strain, the requirement behaves as a recessive in the offspring. A more detailed analysis of this is in progress. It may be more than coincidental that an unexpected effect on the phenotype was found in earlier work[2] when the inversion strain was crossed to the Oregon strain.

The requirement for adenine can be overcome or partially so by altering

the type of medium used. Apparently several ingredients manifest an influence upon the requirement. The role of these various factors is under study. Because of their influence, it was necessary to use the present medium in order to demonstrate an "all-or-none" requirement, even though media supporting more nearly optimal growth are available.[5]

Even though the genetic analysis wants further elucidation, the fact that a multicellular animal has been found with a purine requirement offers the first opportunity for a detailed analysis of purine synthesis in the higher animals under chemically controlled conditions and in the absence of micro-organisms. Such a study is in progress.

* The work was done under a Grant-in-Aid from the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council. The research was aided by a Grant-in-Aid from the Sigma Xi Research Fund.

† The present address of Mr. Ellis is Department of Animal Genetics, University of Edinburgh, Edinburgh, Scotland.

[1] Schultz, Jack, St. Lawrence, P., and Newmeyer, D., *Anat. Rec.*, **96,** 540 (1946) (Abstr.).

[2] Hinton, Taylor, *Am. Nat.*, **83,** 69–94 (1949).

[3] Hinton, Taylor, *Genetics*, **35,** 188–205 (1950).

[4] Schultz, Jack, Caspersson, T., and Aquilonius, L., PROC. NATL. ACAD. SCI., **26,** 515–523 (1940).

[5] Hinton, Taylor, Noyes, D. T., and Ellis, John, *Physiol. Zool.* (in press).

[6] Kidder, G. W., and Dewey, V. C. (unpublished).

[7] Hinton, Taylor, and Ellis, John, *Rec. Gen. Soc. Am.*, **19,** 104–105 (1950) (Abstr.).

[8] Kidder, G. W., and Dewey, V. C., *Arch. Biochem.*, **20,** 433–443 (1949).

---

## *ABSENCE OF CYTOSINE IN BACTERIOPHAGE $T_2$**

BY ALFRED MARSHAK

N.Y.U. COLLEGE OF MEDICINE AND DIVISION OF TUBERCULOSIS, U. S. PUBLIC HEALTH SERVICE

Communicated by Barbara McClintock, February 27, 1951

The common supposition that desoxyribonucleic acid (DNA) is an essential constituent of genic material has found some measure of support in the observation that the relative proportions of the nitrogenous bases are very similar if not identical in DNA's obtained from a wide variety of organisms.[1, 2, 3] Where significant deviations have been found, the possibility of contamination of the DNA could not be ruled out.[4] Observations on the time sequence of the synthesis of DNA and of viable bacteriophage in *E. coli* have been suggested as evidence in support of the supposed genic role of DNA.[5] It is, therefore, disturbing both to the generalization concerning the uniformity of composition of DNA's of diverse origins and to the

assumption that it is an integral genic component to find that the DNA of the *E. coli* bacteriophage $T_2$ is almost if not entirely devoid of cytosine.

The bacteriophage was grown on *Escherichia coli* strain B by the agar layer method of Adams.[6] The agar containing the bacteriophage was extracted twice by blending with 150 ml. of 0.85% NaCl in a Waring blendor and then subjected to centrifugation at 2850 g. for 20 minutes. The supernatant fluid was then centrifuged at 50,000 g. in a vacuum centrifuge for 30 minutes and the sediment resuspended by blending in 0.85% NaCl and filtering through Whatman #41H paper. This cycle was then repeated three times. An aliquot for assay was taken just before the last centrifugation at 50,000 g. The pellet from the last centrifugation was suspended in 20 ml. distilled water, lyophilized and the dry product ground to a powder. Approximately 10 mg. of the powder was hydrolyzed in perchloric acid, and subjected to paper chromatography for separation of the purines and pyrimidines which were assayed spectroscopically.[1, 2] Two other samples of bacteriophage, $T_2$ and $T_{2r}$, obtained through the courtesy of Dr. Seymour Cohen, were analyzed in the same way. The latter two samples had been dialyzed against distilled water for 18 hours before lyophilization. The results are given in the following table.

MOLAR RATIOS

| | ADENINE | GUANINE | THYMINE |
|---|---|---|---|
| $T_2$ (A. M.) | 1.00 ± 0.02 | 0.49 ± 0.02 | 0.90 ± 0.03 |
| $T_2$ (S. C.) | 1.00 ± 0.02 | 0.52 ± 0.03 | 0.96 ± 0.02 |
| $T_{2r}$ (S. C.) | 1.00 ± 0.05 | 0.48 ± 0.03 | 0.95 ± 0.04 |

Four determinations were made of each base in all cases except for thymine where eight determinations were made. The deviations listed are the mean deviations. There are no significant differences in the molar ratios of $T_2$ and the mutant strain $T_{2r}$.

No spots were visible in the chromatograms at the positions for either cytosine or uracil and spectroscopic examination of eluates from these areas showed no absorption throughout the spectral region where these substances absorb. Under the conditions used a spot containing 0.2 $\mu$g. may be detected on the chromatogram and also in the eluate. Since the chromatograms where cytosine would have appeared contained 52 $\mu$g. of adenine, guanine and thymine, any cytosine which might be present must have been less than 0.4% of the total base content. Similar calculations for uracil show that it could not have been present in amounts greater than 0.3%.

Complete absorption curves were taken of all eluates. With adenine and thymine no deviation from the standard curves greater than 2% was observed. In the case of guanine, however, there was a systematic deviation from the standard curve in the region between 270 and 290 m$\mu$ as shown in figure 1.

Since 5-methyl-cytosine has an absorption maximum at 283 m$\mu$ in the region where the deviation from the standard guanine curve was observed, the chromatographic behavior of 5-methyl-cytosine was examined. In the butanol-HCl solvent system used to obtain separation of uracil and thy-

mine, 5-methyl-cytosine moved only very slightly (2 cm.) beyond the starting point. In the system butanol–$NH_4OH$, 5-methyl-cytosine occupied a position between cytosine and adenine, i.e., well removed from guanine which was 5 cm. away. It is very unlikely, therefore, that the observed deviation from the standard guanine curve can be at

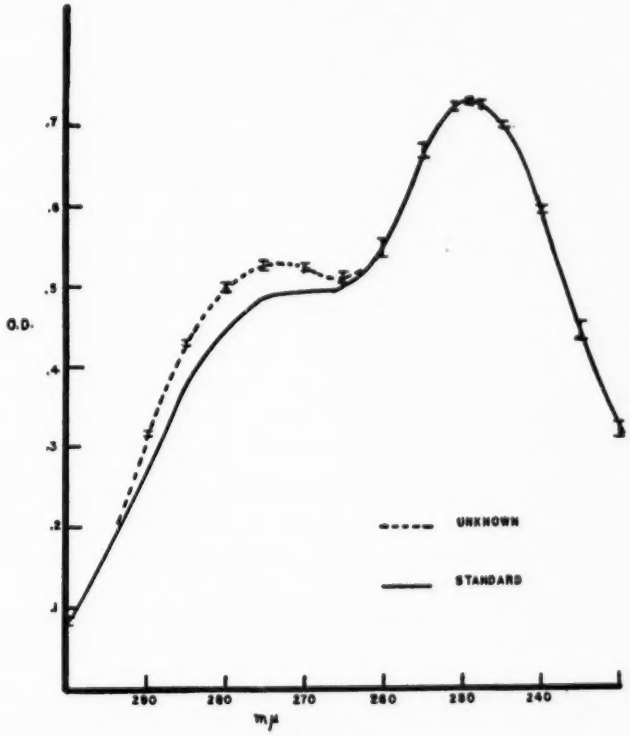



FIGURE 1

Ordinates = optical density
Abscissae = wave length in millimicrons
Horizontal bars indicate the maximum spread of observed values.

tributed to contamination of the unknown with 5-methyl-cytosine. The significance of the unusually high absorption in the 270–290 mμ region of the "guanine" curve remains undetermined. As indicated above, the perfect fit of the absorption curves for the thymine and adenine-containing eluates ruled out the possibility of their containing cytosine or 5-methyl-cytosine. It could be concluded therefore that both of these compounds were not present in the dried products prepared from bacteriophage.

Where the bacteriophage had been dialyzed against distilled water for 18 hours, it seemed possible that cytidylic and uridylic acids might have been preferentially split out and removed, but this could not have been the case where the suspension of bacteriophage was lyophilized directly after isolation. Evidently, then, neither cytosine, 5-methyl-cytosine nor uracil were present in the viable bacteriophage. Since DNA is the major if not the sole nucleic acid of $T_2$,[5] the observations represent an instance where DNA is devoid of cytosine and 5-methyl-cytosine in detectable amounts.

Absence of cytosine is not a general characteristic of the *E. coli* bacteriophages. Wyatt has found appreciable quantities in $T_5$ and preliminary investigations by the writer show considerable amounts in $T_3$. It has been claimed that there are significant differences in the base content of DNA from different mammalian species,[7] although the data actually presented seem to leave this conclusion open to question. A clear difference in composition between the DNA of some bacteria and that of other organisms has been described in which the bacterial product is characterized by an excess of guanine and cytosine as compared with the other bases.[7, 8] However a large excess of these two bases was found in analyses of the total purine and pyrimidine content of bacteria which had a very low DNA as well as a high DNA content.[1, 2] Therefore before accepting the generalization that there are two types of DNA as suggested, it seems necessary that any significant contamination of the microbial DNA used for analysis should be rigidly excluded. However whether or not either of the generalizations mentioned above are accepted, it is evident from the present results that DNA may exist in a form containing no cytosine. An analogous situation has been found in the case of PNA of heart muscle cytoplasm which has no uracil.[9]

In connection with the problem of the synthesis of new phage it should be pointed out that the DNA of the host cell, *E. coli* strain B, has a high cytosine content.[10] Kozloff, *et al.*, have concluded from studies of $N^{15}$ and $P^{32}$ turnover that the bacterial DNA is utilized for the synthesis of new virus particles, and that the major portion of the bacterial N contribution is found in the phage DNA.[11] If it is indeed true that the phage utilizes the host DNA it must do so in a manner which rejects, eliminates or transforms the cytosine which the latter contains. Since nuclear PNA, in contrast to cytoplasmic PNA, has a very low cytosine content and also since in other species the nuclear PNA appears to be the precursor of DNA,[9, 12] the results suggest that the former rather than the latter may be involved in phage synthesis.

It has been suggested or implied that DNA has a genic function and that mutations occur by virtue of small changes at the molecular level such as, for example, alterations in the number or arrangement of particular groupings in the molecule.[7, 13, 14] The absence of cytosine and of 5-methyl-

cytosine in $T_2$ presents a serious obstacle to the acceptance of the supposed genic role of DNA, at least as it is usually conceived. $T_2$ is not exceptional among the *E. coli* phages in the sense that it carries out the usual processes of adsorption by the bacterial cell and proliferation within it. If the DNA of $T_2$ can perform its usual function even though it lacks one of the components (cytosine) usually present as a major constituent of DNA from other sources, it must be concluded that this function is not necessarily dependent upon specific structure as determined by a particular configuration of groups of atoms within the molecule but rather that DNA may exert its influence in whatever role it plays by some more general property.

* Supported by a grant from the U. S. Atomic Energy Commission under Contract No. AT (30-1)-1038.

[1] Marshak, A., and Vogel, H., *Federation Proc.*, **9**, 85–86 (1950).

[2] Marshak, A., and Vogel, H., *J. Biol. Chem.* (1951), in press.

[3] Daly, M. M., Allfrey, V. G., and Mirsky, A. E., *J. Gen. Physiol.*, **33**, 497–510 (1950).

[4] Wyatt, G. R., *Biochem. J.*, **47**, VII (1950).

[5] Cohen, S. S., and Arbogast, R., *J. Exp. Med.*, **91**, 607–618 (1950).

[6] Adams, M., *Methods in Medical Research*, Vol. 2, 1–73 (1950), Year Book Publishers, Inc., Chicago.

[7] Chargaff, E., *Experientia*, **6**, 201–209 (1950).

[8] Chargaff, E., Zamenhof, S., Brawerman, G., and Kerin, L., *J. Am. Chem. Soc.*, **72**, 3825 (1950).

[9] Marshak, A., *J. Biol. Chem.*, in press.

[10] Marshak, A., Unpublished data.

[11] Kozloff, L. W., Knowlton, K., Putnam, F. W., and Evans, E. A., Jr., *Federation Proc.*, **9**, 192 (1950).

[12] Marshak, A., *J. Cellular Comp. Physiol.*, **32**, 381–406 (1948).

[13] Muller, H. J., *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 151–167 (1941).

[14] Stern, Kurt, *Yale J. Biol. and Med.*, **19**, 937–949 (1946–1947).

---

## *ON THE GENERATING FUNCTIONS OF TOTALLY POSITIVE SEQUENCES**

BY MICHAEL AISSEN, ALBERT EDREI, I. J. SCHOENBERG, AND ANNE WHITNEY

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PENNSYLVANIA, AND THE DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SASKATCHEWAN

Communicated by Oswald Veblen, March 17, 1951

A real matrix, finite or infinite, is called *totally positive* if and only if all its minors, of all orders 1, 2, ..., are non-negative. An infinite real sequence $\{a_n\}$, $(-\infty < n < \infty)$, is called *totally positive* if and only if the corresponding 4-way infinite matrix

$$\|a_{i-k}\| = \begin{Vmatrix} \vdots & \vdots & \vdots \\ \ldots a_0 & a_{-1} & a_{-2}\ldots \\ \ldots a_1 & a_0 & a_{-1}\ldots \\ \ldots a_2 & a_1 & a_0 \ \ldots \\ \vdots & \vdots & \vdots \end{Vmatrix},$$

is totally positive. In the present note we are concerned only with sequences such that $a_n = 0$ if $n < 0$, and $a_0 = 1$. Accordingly we shall say that the ordinary sequence $a_0, a_1, a_2, \ldots$ $(a_0 = 1)$ is *totally positive* meaning thereby that the sequence $\ldots 0, 0, 0, a_0, a_1, a_2, \ldots$ is totally positive; it is readily seen that this is the case if and only if the 2-way infinite matrix

$$A = \|a_{i-k}\| = \begin{Vmatrix} a_0 & 0 & 0\ldots \\ a_1 & a_0 & 0\ldots \\ a_2 & a_1 & a_0\ldots \\ \vdots & & \vdots \end{Vmatrix} \qquad (i, k = 1, 2, 3, \ldots), \quad (1)$$

is totally positive.[1] The problem solved in this note is a characterization of the generating functions

$$f(z) = a_0 + a_1 z + a_2 z^2 + \ldots \qquad (2)$$

of totally positive sequences (Theorem 4 below). As the detailed results are going to appear in three notes by the four authors in different combinations, the contributions of each will be made clear in the present note.

The following obvious remarks are already in the literature:[2] The class of generating functions of totally positive sequences is closed with respect to multiplication, because to the multiplication of generating functions corresponds the multiplication of the corresponding matrices (1) while the product of two totally positive matrices is clearly totally positive. The class of generating functions is also obviously closed with respect to passage to the limit. These remarks, together with the observation that the three functions

$$1 + \alpha z, \qquad \frac{1}{1 - \beta z}, \qquad e^{\gamma z} \qquad (\alpha \geqq 0, \beta \geqq 0, \gamma \geqq 0),$$

generate totally positive sequences, lead to the following result:

*1. The functions $f(z)$ of the form*

$$f(z) = e^{\gamma z} \cdot \frac{\prod_{\nu=1}^{\infty} (1 + \alpha_\nu z)}{\prod_{\nu=1}^{\infty} (1 - \beta_\nu z)} \qquad (3)$$

$$(\gamma \geqq 0, \alpha_\nu \geqq 0, \beta_\nu \geqq 0, \textstyle\sum \alpha_\nu, \sum \beta_\nu \text{ convergent}),$$

*generate totally positive sequences.*

In a paper presented at the International Congress of Mathematicians,

September, 1950. Aissen, Schoenberg and Whitney have proved a partial converse of Theorem 1 which is as follows:

*2. If $a_0, a_1, a_2, \ldots$ is a totally positive sequence, then its generating function $f(z)$ is regular in a neighborhood of the origin and may be extended into the whole finite plane, being a meromorphic function of the following special form*

$$F(z) = e^{g(z)} \cdot \frac{\prod_1^\infty (1 + \alpha_\nu z)}{\prod_1^\infty (1 - \beta_\nu z)} \tag{4}$$

$$(\alpha_\nu \geqq 0,\ \beta_\nu \geqq 0,\ \textstyle\sum \alpha_\nu,\ \sum \beta_\nu \text{ convergent}),$$

*where $g(z)$ is an entire function such that the exponential factor $\exp\{g(z)\}$ will by itself also generate a totally positive sequence.*

This is a problem of the general type discussed by Hadamard in his classical dissertation "Essai sur l'étude des fonctions données par leur développement de Taylor," of 1892.[3] Hadamard's work furnished some of the main ideas used in obtaining (4), adapted to our particular situation which gives the precise information concerning the nature of the zeros and poles implied by (4). An important tool is a criterion by Anne Whitney allowing one to decide if an ordinary $m$ by $n$ matrix is totally positive by reducing easily the problem to the similar one for a $m - 1$ by $n$ matrix (reduction theorem).[4]

The main reason, however, why this note is being written is that Albert Edrei succeeded in proving recently the following result which we have conjectured for some time:

*3. If $g(z) = c_1 z + c_2 z^2 + \ldots$ is entire and such that $e^{g(z)}$ generates a totally positive sequence, then we necessarily have that $c_2 = c_3 = \ldots = 0$, so that $g(z) = c_1 z$, $(c_1 \geqq 0)$.*

This result is obtained by combining Theorem 2 with a refinement of the Picard theorem for entire functions due to R. Nevanlinna. Applying Theorem 3 to the conclusion of Theorem 2 we learn that $g(z)$ appearing in (4) must reduce to the form $g(z) = \gamma z$ $(\gamma \geqq 0)$. We may thus state the following final theorem.[5]

*4. The sequence $a_0, a_1, a_2, \ldots$, $(a_0 = 1)$, is totally positive if and only if it is generated by a function $f(z)$ of the form*

$$f(z) = e^{\gamma z} \frac{\prod_1^\infty (1 + \alpha_\nu z)}{\prod_1^\infty (1 - \beta_\nu z)}. \tag{5}$$

$$(\gamma \geqq 0,\ \alpha_\nu \geqq 0,\ \beta_\nu \geqq 0,\ \textstyle\sum \alpha_\nu,\ \sum \beta_\nu \text{ convergent}).$$

As a special case we have the following theorem

*5. Let the function* $f(z) = a_0 + a_1 z + \ldots$ $(a_0 = 1)$ *be entire; then the sequence* $a_0, a_1, a_2, \ldots$ *is totally positive if and only if* $f(z)$ *is of the form*

$$f(z) = e^{\gamma z} \prod_1^\infty (1 + \alpha_\nu z) \qquad (\gamma \geqq 0,\ \alpha_\nu \geqq 0,\ \textstyle\sum \alpha_\nu \text{ convergent}). \quad (6)$$

A very special case of Theorem 5 is

*6. The real polynomial* $P(z) = a_0 + a_1 z + \ldots + a_n z^n (a_0 = 1)$ *has only real and negative zeros if and only if the sequence* $a_0, a_1, \ldots, a_n, 0, 0, 0, \ldots$ *is totally positive.*

As an application of the main Theorem 4 we mention the following theorem

*7. If* $f(z) = a_0 + a_1 z + a_2 z^2 + \ldots$ *is the expansion of a function of the form (5) and* $k$ *and* $l$ *are integers such that* $0 \leqq k < l$, *then also*

$$f_{k,l}(z) = a_k + a_{k+l} z + a_{k+2l} z^2 + \ldots + a_{k+nl} z^n + \ldots$$

*is the expansion of a function of the form (5) multiplied by the constant* $a_k$.[6]

For indeed, by the direct part of Theorem 4 the matrix $\|a_{i-j}\|$ $(a_n = 0$ if $n < 0)$ is totally positive, hence so is the matrix $\|a_{k+l(i-j)}\|$, being a submatrix of $\|a_{i-j}\|$. Thus $a_k, a_{k+l}, a_{k+2l}, \ldots$ is a totally positive sequence and therefore $f_{k,l}(z)$ is of the form (5) by the converse part of Theorem 4. An equally evident application of Theorem 4 is the following theorem

*8. If* $\{f_n(z)\}$ *is an infinite sequence of functions of the form (5) and* $\lim f_n(z) = f(z)$, *uniformly in a neighborhood of the origin, then also the limit function* $f(z)$ *is of the form (5).*

These results reveal with clarity the relation between the class of entire functions of the form (6) and its subclass of polynomials with only negative zeros. This is shown by the ease with which we may now derive the following classical theorem of Laguerre and Polya:[7]

*9. If* $\{P_n(z)\}$ *is an infinite sequence of polynomials with only negative zeros and constant term unity, and if* $\lim P_n(z) = f(z)$, *uniformly in every finite domain, then the limit function* $f(z)$ *is of the form (6).*

Indeed, each $\{P_n(z)\}$ is of the form (5) and therefore so is the limit function $f(z)$ by Theorem 8; since $f(z)$ is entire, it must reduce to the form (6).

* The subject of the present note is part of a project sponsored by the Office of Naval Research at the University of Pennsylvania.

[1] A related concept was first introduced by M. Fekete in 1912; Fekete calls the function (2) *n-times positive* if all those minors of the matrix (1) are positive which are of an order $= 1, 2, \ldots, n + 1$, and do not vanish identically for formal reasons (because of their vanishing elements). See Fekete, M., and Polya, G., "Über ein Problem von Laguerre," *Rendiconti di Palermo*, **34**, 1–32 (1912), especially pages 3–7.

[2] Schoenberg, I. J., "Some Analytical Aspects of the Problem of Smoothing," *Courant Anniversary Volume*, New York, 1948, pp. 351–370, especially §2.

[3] Hadamard, J., *Jour. de Math.*, 4ᵉ série, **8** (1892).

[4] In a paper presented by Anne Whitney at the International Congress of Mathematicians, September, 1950.

[5] This is a discrete analog of a problem solved by Schoenberg; see the paper "On Totally Positive Functions, Laplace Integrals and Entire Functions of the Laguerre-Polya-Schur Type," PROC. NATL. ACAD. SCI., **33,** 11–17 (1947).

[6] For the special case when $f(z)$ is entire, hence of the form (6), this Theorem 7 is due to Polya; see Polya, G., "Über einen Satz von Laguerre," *Jahresb. der Deutschen Mathematiker-Vereinigung,* **38,** 161–168 (1929), especially pages 166–168.

[7] Laguerre, "Sur les fonctions du genre zéro et du genre un," *Oeuvres de Laguerre,* Vol. 1, Paris, 1898, pp. 174–177, and Polya, G., "Über Annäherung durch Polynome mit lauter reellen Wurzeln," *Rendiconti di Palermo,* **36,** 1–17 (1913).

---

# *COHOMOLOGY THEORY OF ABELIAN GROUPS AND HOMOTOPY THEORY III*

BY SAMUEL EILENBERG* AND SAUNDERS MacLANE

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND THE UNIVERSITY OF CHICAGO

Communicated March 31, 1951

*1. Introduction.*—For each abelian group $\Pi$ and for each integer $n$ we consider the complexes $K(\Pi, n)$ described in Note I[1] and the suspension maps $S{:}K(\Pi, n) \to K(\Pi, n+1)$ raising the dimensions by 1. We recall that $K(\Pi, n)$ gives the singular homology and cohomology theory for any arcwise connected space $X$ in which only the $n$th homotopy group $\pi_n(X) = \Pi$ is non-trivial ($n > 1$). We show now that the augmented complex $K(\Pi, n)$ is, after a change in dimension, chain equivalent to the complex $A^{n-1}(\Pi)$ which gives the "abelian" homology theory of $\Pi$ in level $n-1$. This proves the main conjecture (26) of Note II.[1]

*2. Products.*—The complex $K(\Pi, n)$ is a complete semisimplicial complex.[2] This means the following. If $[q]$ denotes the set of integers $i$ with $0 \leqq i \leqq q$ and $\alpha : [p] \to [q]$ any weakly monotone map, each $q$-simplex $\sigma_q$ of $K(\Pi, n)$ determines a $p$-simplex $\sigma_q\alpha$, with appropriate formal properties. In particular, the $i$th face of $\sigma_q$ is the $(q-1)$-simplex $F_i\sigma_q = \sigma_q\epsilon^i$, where $\epsilon^i : [q-1] \to [q]$ is the monotone map covering all of $[q]$ except the integer $i$. We also use the degenerate simplices $N_i\sigma_q = \sigma_q\eta^i$, defined for $0 \leqq i \leqq q$ by the maps $\eta^i = \eta_q{}^i$: $[q+1] \to [q]$ with $\eta^i(j) = j$, for $j \leqq i$ and $\eta^i(j) = j-1$ for $i < j$. We denote by $K'(\Pi, n)$ the quotient of $K(\Pi, n)$ by the subcomplex consisting of all degenerate cells and the single 0-cell. If $K_a(\Pi, n)$ is the complex $K(\Pi, n)$ augmented in dimension $-1$ by the group of integers, then the natural map $K_a(\Pi, n) \to K'(\Pi, n)$ is a chain equivalence, by a known normalization theorem.[2] The suspension map $S{:}K'(\Pi, n) \to K'(\Pi, n+1)$ is still valid.

Because of the addition in $\Pi$ we may construct for any two $q$-simplices $\sigma$ and $\tau$ in $K(\Pi, n)$ a new $q$-simplex $\sigma \oplus \tau$, their *internal sum*, defined as the sum of $\sigma$ and $\tau$, considered (Note I) as elements of $Z^n(\Delta_q, \Pi)$. Using this sum, we define two products $*$ and $\downarrow$, as follows. For any non-negative integers $p$ and $q$ consider all partitions of the set $[p + q - 2]$ into two disjoint sets $a$ and $b$, with $q - 1$ and $p$ elements, respectively. Write the integers in the set $a$ in descending order as $a = \{a_{q-1}, \ldots, a_1\}$ and set

$$N_a = N_{a_{q-1}} \ldots N_{a_1}, \qquad \epsilon(a) = \sum_{j=1}^{q-1} a_j - (j - 1). \tag{1}$$

For simplices $\sigma_p$ and $\tau_q$ of dimensions $p$ and $q$, respectively, in $K(\Pi, n)$ a chain $\sigma_p \downarrow \tau_q$ of dimension $p+q$ is now defined as

$$\sigma_p \downarrow \tau_q = \sum (-1)^{\epsilon(b)} [N_a N_p \sigma_p \oplus N_b \tau_q], \tag{2}$$

with the sum taken over all partitions $a$, $b$ as above. The formula

$$\sigma_p * \tau_q = \sigma_p \downarrow \tau_q + (-1)^{pq} \tau_q \downarrow \sigma_p \tag{3}$$

now yields a product of excess zero, in the sense of Note II. It has a "geometric" interpretation in terms of the triangulation of the cartesian product $\Delta_p \times \Delta_q$ of standard affine simplices. Both these products carry over to $K'(\Pi, n)$, and $*$ is still a product of excess zero there.

*3. The Main Transformation.*—For each cell complex $L$ with a product of excess 0, we have defined in Note II, by means of the so-called "bar construction," a complex $\mathfrak{B}(L)$ with a product of excess 1, containing $L$ as a subcomplex. We shall denote by $\mathfrak{B}^+(L)$ the complex obtained from $\mathfrak{B}(L)$ by raising all dimensions one unit. Then in $\mathfrak{B}^+(L)$ the product has excess 0, while the inclusion map $L \to \mathfrak{B}(L)$ becomes a map $S_B: L \to \mathfrak{B}^+(L)$ raising the dimension by 1.

We now define a chain transformation

$$f: \mathfrak{B}^+(K'(\Pi, n)) \to K'(\Pi, n + 1). \tag{4}$$

Specifically, if $\sigma$ is a $p$-simplex of $K'(\Pi, n)$, and hence a $p + 1$ simplex of $\mathfrak{B}^+(K'(\Pi, n + 1))$, we set $f(\sigma) = S(\sigma)$. Any other cell of the complex on the left has one or more bars, and thus may be written as $[\rho \mid \sigma]$, where $\rho$ is a cell with one fewer bar, and $\sigma$ a cell of $K'(\Pi, n)$. We thus define $f$ by induction on the number of bars as

$$f[\rho \mid \sigma] = f\rho \downarrow f\sigma, \qquad f\sigma = S(\sigma). \tag{5}$$

(The product $\downarrow$ is not associative.)

THEOREM 1. *The map $f$ of (4) is a chain equivalence, with the additional properties*

$$f(\rho_1 * \rho_2) = f(\rho_1) * f(\rho_2), \tag{6a}$$

$$fS_B = S. \tag{6b}$$

The proof depends upon a representation of $K(\Pi, n+1)$ in terms of $K(\Pi, n)$. Specifically, any $q+1$ cell of $K(\Pi, n+1)$ can be represented uniquely as a $(q+1)$-tuple

$$\tau_{q+1} = \langle \sigma_q, \sigma_{q-1}, \ldots, \sigma_0 \rangle, \tag{7}$$

where each $\sigma_i$ is an $i$-cell of $K(\Pi, n)$. The semisimplicial structure of $K(\Pi, n+1)$ is determined by giving the operator for the $i$th face $F_i$, as

$$\begin{aligned} F_0\tau_{q+1} &= \langle F_0\sigma_q, \ldots, F_0\sigma_1 \rangle, \\ F_i\tau_{q+1} &= \langle F_i\sigma_q, \ldots, F_i\sigma_{i+1}, F_i\sigma_i \oplus \sigma_{i-1}, \sigma_{i-2}, \ldots, \sigma_0 \rangle, \\ F_{q+1}\tau_{q+1} &= \langle \sigma_{q-1}, \ldots, \sigma_0 \rangle, \end{aligned}$$

for $0 < i < q+1$, and the operator for the $i$th "degenerate" simplex as

$$N_i\tau_{q+1} = \langle N_i\sigma_q, \ldots, N_i\sigma_i, 0_i, \sigma_{i-1}, \ldots, \sigma_0 \rangle,$$

with $0 \leqq i \leqq q+1$, where $0_i$ denotes the cell (cocycle) which is identically zero.

By means of this representation, we introduce in $K'(\Pi, n+1)$ a family of subcomplexes $K_k'$, spanned by those cells $\tau$ with at most $k$ non-zero entries in (7). In $\mathfrak{B}^+(K(\Pi, n)\mathrm{F})$ we use the family of subcomplexes $\mathfrak{B}_k{}^+$, spanned by the cells with less than $k$ bars ($k = 1, 2, \ldots$). The map $f$ of (4) carries $\mathfrak{B}_k{}^+$ into $K_k'$, and the proof of the Theorem is completed by showing that each of the induced maps for the factor complexes $\mathfrak{B}_{k+1}{}^+/\mathfrak{B}_k{}^+ \to K_{k+1}'/K_k'$ is a chain equivalence.

The proof of Theorem 1 also is independent of the value of the integer $n$, and actually yields a more general theorem as to the equivalence of the processes $\mathfrak{B}^+$ and (7), when applied to a complete semisimplicial complex which has a suitable internal sum and which has only one cell in dimension zero.

4. *The Abelian Homology Theory.*—Let $A(\Pi, 1)$ denote the complex $K(\Pi, 1)$ with the single 0-simplex removed. We define two sequences of complexes $A(\Pi, n)$, $A'(\Pi, n)$, as follows

$$A(\Pi, n+1) = \mathfrak{B}^+(A(\Pi, n)), \tag{8}$$

$$A'(\Pi, 1) = K'(\Pi, 1), \qquad A'(\Pi, n+1) = \mathfrak{B}^+(A'(\Pi, n)). \tag{9}$$

These complexes all have products (of excess 0) and maps $S_A : A(\Pi, n) \to A(\Pi, n+1)$ and $S_A' : A'(\Pi, n) \to A'(\Pi, n+1)$, raising the dimension by 1. If in $A(\Pi, n)$ all dimensions are lowered by $n-1$, there results the complex $A^{n-1}(\Pi)$ of Note II.[3] In virtue of the normalization theorem quoted above, there is a chain equivalence $\eta : A(\Pi, 1) \to A'(\Pi, 1)$, hence, by the iteration process described in Theorem 2 of Note II[4] there are chain equivalences $\eta_n : A(\Pi, n) \to A'(\Pi, n)$, which preserve the products and commute with the suspensions $S_A$. Using Theorem 1 above and Theorem 2 of Note II we obtain

THEOREM 2. *Iteration on the map f of (4) yields maps*

$$f_n: A'(\Pi, n) \to K'(\Pi, n)$$

*which are chain equivalences, satisfy* $f_n(\sigma * \tau) = f_n(\sigma) * f_n(\tau)$, *and yield a commutative diagram*

$$\begin{array}{ccccc} \cdots \to A'(\Pi, n) & \xrightarrow{S^*A} & A'(\Pi, n+1) & \to & \cdots \\ \downarrow f_n & & \downarrow f_{n+1} & & \\ \cdots \to K'(\Pi, n) & \xrightarrow{S} & K'(\Pi, n+1) & \to & \cdots \end{array}$$

This result has as a corollary the suspension Theorem (Theorem 1) of Note I. It also proves that the conjecture expressed in formula (26) of Note II is correct.

The homology and cohomology groups of $K(\Pi, n)$ over a coefficient group $G$ are denoted by $H_q(\Pi, n; G)$ and $H^q(\Pi, n; G)$. The groups derived from the complexes $A(\Pi, n)$ have, as shown in Note II, a direct connection with the so-called homology theory of abelian groups.[5] Theorem 2 yields isomorphisms

$$H_q(\Pi, n; G) \cong H_q(A(\Pi, n); G), \quad q > 0,$$
$$H^q(\Pi, n; G) \cong H^q(A(\Pi, n); G), \quad q > 0,$$

without the use of the intermediate cubical complexes $Q_t(\Pi)$ of Note I. When $\Pi$ and $G$ are finitely generated these groups are also finitely generated (and finitely computable), as observed by Serre.[6]

[*] John Simon Guggenheim Memorial Fellow.

[1] Eilenberg, S., and MacLane, S., PROC. NATL. ACAD. SCI., **36**, 443–447 and 657–663 (1950); subsequently referred to as Note I and Note II.

[2] Eilenberg, S., and Zilber, J. A., *Ann. Math.*, **51**, 499–513 (1950).

[3] In Notes I and II we failed to distinguish between $K(\Pi, 1)$ and $K(\Pi, 1)$ with the zero cell omitted.

[4] In Note II (p. 659) in defining a reduction $f: L \to L'$, we needlessly required that $f$ map $L$ onto $L'$.

[5] We take this opportunity to correct the following computational error, called to our attention by H. Cartan, in Note II. On p. 662, instead of $A_{\Pi}(J) = (2) + (2) + (2)$, read $A_{\Pi}(J) = (2) + (2)$.

[6] Serre, J. P., *Compt. rend.*, **232**, 142–144 (1951).

# GENERALIZED RIEMANN SPACES

BY LUTHER P. EISENHART

FINE HALL, PRINCETON, NEW JERSEY

Communicated March 17, 1951

A space of coordinates $x^i (i = 1, \ldots, n)$ with which is associated a non-symmetric tensor $g_{ij}$ is termed a generalized Riemann space. The components $g_{ij}$ and $g'_{\alpha\beta}$ in two coordinate systems $x^i$ and $x'^\alpha$ are related by the equation

$$g'_{\alpha\beta} = g_{ij} \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial x}{\partial x'^\beta}. \tag{1}$$

The summation convention respecting dummy indices, such as $i$ and $j$, is employed here and throughout this paper.

Differentiating equation (1) with respect to $x'^\gamma$ we have

$$\frac{\partial g'_{\alpha\beta}}{\partial x'^\gamma} = \frac{\partial g_{ij}}{\partial x^k} \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial x^j}{\partial x'^\beta} \frac{\partial x^k}{\partial x'^\gamma} + g_{ij} \left( \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial^2 x^j}{\partial x'^\beta \partial x'^\gamma} + \frac{\partial^2 x^i}{\partial x'^\alpha \partial x'^\gamma} \frac{\partial x^j}{\partial x'^\beta} \right). \tag{2}$$

The first of the following equations is obtained by interchanging the indices $\alpha$ and $\gamma$ throughout and the dummy indices $i$ and $k$ in the first term of the right-hand member; the second by interchanging the indices $\beta$ and $\gamma$ and the dummy indices $j$ and $k$:

$$\left.\begin{aligned} \frac{\partial g'_{\gamma\beta}}{\partial x'^\alpha} &= \frac{\partial g_{kj}}{\partial x^i} \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial x^j}{\partial x'^\beta} \frac{\partial x^k}{\partial x'^\gamma} + g_{ij} \left( \frac{\partial x^i}{\partial x'^\gamma} \frac{\partial^2 x^j}{\partial x'^\beta \partial x'^\alpha} + \frac{\partial^2 x^i}{\partial x'^\alpha \partial x'^\gamma} \frac{\partial x^j}{\partial x'^\beta} \right), \\ \frac{\partial g'_{\alpha\gamma}}{\partial x'^\beta} &= \frac{\partial g_{ik}}{\partial x^j} \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial x^j}{\partial x'^\beta} \frac{\partial x^k}{\partial x'^\gamma} + g_{ij} \left( \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial^2 x^j}{\partial x'^\beta \partial x'^\gamma} + \frac{\partial^2 x^i}{\partial x'^\alpha \partial x'^\beta} \frac{\partial x^j}{\partial x'^\gamma} \right). \end{aligned}\right\} \tag{3}$$

We introduce the notation

$$\Delta_{ijk} = \frac{1}{2} (g_{ik,\,j} + g_{kj,\,i} - g_{ij,\,k}), \tag{4}$$

where throughout this paper a quantity followed by a comma and index indicates the derivative of the quantity with respect to the coordinate with that index; thus $g_{ik,\,j} \equiv \dfrac{\partial g_{ik}}{\partial x^j}$. By means of this notation one-half the sum of equations (3) minus one-half of equation (2) may be written

$$\Delta'_{\alpha\beta\gamma} = \Delta_{ijk} \frac{\partial x^i}{\partial x'^\alpha} \frac{\partial x^j}{\partial x'^\beta} \frac{\partial x^k}{\partial x'^\gamma} + g_{ij} \frac{\partial x^i}{\partial x'^\gamma} \frac{\partial^2 x^j}{\partial x'^\alpha \partial x'^\beta} \tag{5}$$

where $g_{ij}$ denotes the symmetric part of $g_{ij}$.

We specify that the determinant of $g_{ij}$ is different from zero, and define quantities $g^{ij}$ by the equation

$$g_{ij}g^{ik} = \delta_j^{\ k},$$

where $\delta_j^{\ k}$ are the Kronecker deltas. By means of the notation

$$\Delta_{ij}^{\ \ h} = g^{hk}\Delta_{ijk}, \qquad \Delta_{ijk} = g_{hk}\Delta_{ij}^{\ \ h}, \tag{6}$$

equations (5) are equivalent to[1]

$$\frac{\partial^2 x^h}{\partial x'^\alpha \partial x'^\beta} + \Delta_{ij}^{\ \ h}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x}{\partial x'^\beta} = \Delta'_{\alpha\beta}^{\ \ \gamma}\frac{\partial x^h}{\partial x'^\gamma} \tag{7}$$

From (4) one obtains

$$\Delta_{ikj} + \Delta_{jki} = g_{ij,\,k}$$

which by (6) may be written

$$g_{ij,\,k} = g_{hj}\Delta_{ik}^{\ \ h} + g_{hi}\Delta_{jk}^{\ \ h}. \tag{8}$$

Also

$$\Delta_{kij} + \Delta_{kji} = g_{ij,\,k},$$

that is,

$$g_{ij,\,k} = g_{hj}\Delta_{ki}^{\ \ h} + g_{hi}\Delta_{kj}^{\ \ h}. \tag{9}$$

One-half the sum of equations (8) and (9) is

$$g_{ij,\,k} = g_{ij}\Delta_{ik}^{\ \ h} + g_{hi}\Delta_{jk}^{\ \ h}, \tag{10}$$

where $\Delta_{ik}^{\ \ h}$ denotes the symmetric part of $\Delta_{ik}^{\ \ h}$. From (10) it follows that

$$\Delta_{ij}^{\ \ h} = \left\{\begin{matrix} h \\ ij \end{matrix}\right\}, \tag{11}$$

where $\left\{\begin{matrix} h \\ ij \end{matrix}\right\}$ is the Christoffel symbol of the second kind in the quantities $g_{ij}$.[2]

Denoting by $\Delta_{ij}^{\ \ h}$, $\Delta_{ijk}$ and $g_{ij}$ the skew-symmetric parts $\Delta_{ij}^{\ \ h}$, $\Delta_{ijk}$, and $g_{ij}$ respectively, we have from (4)

$$g_{hk}\Delta_{ij}^{\ \ h} = \Delta_{ijk} = \frac{1}{2}(g_{ik,\,j} + g_{kj,\,i} + g_{ji,\,k}). \tag{12}$$

A consequence of (12) is

$$g_{hk}\Delta_{ij}^{\ \ h} = g_{hi}\Delta_{jk}^{\ \ h} = g_{hj}\Delta_{ki}^{\ \ h}, \tag{13}$$

which is in agreement with the result of subtracting equation (9) from equation (8). Also one has from (6) and (13)

$$\Delta_i \equiv \Delta_{ij}^{\ j} = g^{jk}\Delta_{ijk} = g^{jk}\Delta_{jki} = 0. \tag{14}$$

The quantities $\Delta_{ij}^{\ h}$ are coefficients of an affine connection of the space. Any other set $\Gamma_{ij}^{\ h}$ in two coordinate systems $x^i$ and $x'^\alpha$ are related by the equations

$$\frac{\partial^2 x^h}{\partial x'^\alpha \partial x'^\beta} + \Gamma_{ij}^{\ h}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = \Gamma'_{\alpha\beta}{}^{\gamma}\frac{\partial x^h}{\partial x'^\gamma} \tag{15}$$

analogous to equations (7). Subtracting this equation from (7) we have

$$(\Gamma_{ij}^{\ h} - \Delta_{ij}^{\ h})\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = (\Gamma'_{\alpha\beta}{}^{\gamma} - \Delta'_{\alpha\beta}{}^{\gamma})\frac{\partial x^h}{\partial x^\gamma}$$

that is, $a_{ij}^{\ h}$, defined by

$$\Gamma_{ij}^{\ h} - \Delta_{ij}^{\ h} = a_{ij}^{\ h}, \tag{16}$$

are components of a tensor, and for any tensor $a_{ij}^{\ h}$ and $\Delta_{ij}^{\ h}$ one obtains an affine connection.

Interchanging $\alpha$ and $\beta$ in (15) and the dummy indices $i$ and $j$, and subtracting the resulting equation from (15), one obtains

$$2\Gamma_{[ij]}^{\ h}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta} = 2\Gamma'_{\alpha\beta}{}^{\gamma}\frac{\partial x^h}{\partial x'^\gamma}$$

that is, the skew-symmetric part $\Gamma_{[ij]}^{\ h}$ of $\Gamma_{ij}^{\ h}$ is a tensor. Similarly $\Delta_{[ij]}^{\ h}$ is a tensor and by (14) $\Delta_i$ is a zero vector.

We write

$$g_{ij,\,k} = g_{hj}\Delta_{ik}^{\ h} + g_{ih}\Delta_{jk}^{\ h} + a_{ijk} \tag{17}$$

where $a_{ijk}$ are the components of a tensor whose form is to be determined. From (17) we have

$$g_{ij,\,k} = g_{hj}\Delta_{ik}^{\ h} + g_{ih}\Delta_{jk}^{\ h} + \frac{1}{2}(a_{ijk} + a_{jik}).$$

Comparing this equation with equation (8) we note that $a_{ijk}$ is skew-symmetric in the first two indices. Also from (17) one has

$$g_{ij,\,k} = g_{hj}\Delta_{ik}^{\ h} + g_{ih}\Delta_{jk}^{\ h} + a_{ijk} \tag{18}$$

From equations of this form one obtains

$$\frac{1}{2}(g_{ji,\,k} + g_{kj,\,i} + g_{ik,\,j}) = g_{hi}\Delta_{jk}^{\ h} + g_{hj}\Delta_{ki}^{\ h} + g_{hk}\Delta_{ij}^{\ h} - \frac{1}{2}(a_{ijk} + a_{jki} + a_{kij}). \tag{19}$$

This equation is satisfied by

$$a_{ijk} = 2g_{h\underset{\smile}{k}}\Delta_{\underset{\smile}{ij}}^{\ h} - \frac{2}{3}\Delta_{\underset{\smile}{ijk}} \tag{20}$$

in consequence of (12) and (13).

If one adds to the right-hand number of (20) the tensor $b_{ijk}$ skew-symmetric in the first two indices the equation (19) is satisfied provided that

$$b_{ijk} + b_{jki} + b_{kij} = 0.$$

An example of a tensor satisfying this condition is

$$b_{ijk} = (\lambda_{i,\,j} - \lambda_{j,\,i})_{,k}$$

where $\lambda_i$ is any covariant vector.

The condition of integrability of equations (15) is reducible by means of equations of the form (15) to

$$\Gamma_{ijk}^{\ \ h}\frac{\partial x^i}{\partial x'^\alpha}\frac{\partial x^j}{\partial x'^\beta}\frac{\partial x^k}{\partial x'^\gamma} = \Gamma'_{\alpha\beta\ \gamma}^{\ \ \delta}\frac{\partial x^h}{\partial x'^\delta}$$

where

$$\Gamma_{ij,\ k}^{\ \ h} = \Gamma_{ij,\ k}^{\ \ h} - \Gamma_{i\underset{\smile}{k}\ j}^{\ \ h} + \Gamma_{ij}^{\ \ l}\Gamma_{lk}^{\ \ h} - \Gamma_{i\underset{\smile}{k}}^{\ \ l}\Gamma_{lj}^{\ \ h}. \tag{21}$$

Hence $\Gamma_{ij}^{\ \ h}{}_k$ are components of a tensor.

Denoting by $\Gamma_{ij}$ the tensor $\Gamma_{ij}^{\ \ h}{}_h$, one has

$$\Gamma_{ij} = \Gamma_{ij}^{\ \ h}{}_{,h} - \Gamma_{i\underset{\smile}{h}}^{\ \ h}{}_{,j} + \Gamma_{ij}^{\ \ l}\Gamma_{lh}^{\ \ h} - \Gamma_{i\underset{\smile}{h}}^{\ \ l}\Gamma_{lj}^{\ \ h}, \tag{22}$$

The symmetric and skew-symmetric parts of $\Gamma_{ij}$ are

$$\left.\begin{aligned}\Gamma_{\underset{\frown}{ij}} &= \Gamma_{\underset{\frown}{ij}}^{\ \ h}{}_{,h} - \frac{1}{2}(\Gamma_{i\underset{\smile}{h}}^{\ \ h}{}_{,j} + \Gamma_{j\underset{\smile}{h}}^{\ \ h}{}_{,i}) + \Gamma_{\underset{\frown}{ij}}^{\ \ l}\Gamma_{lh}^{\ \ h} - \Gamma_{\underset{\frown}{ih}}^{\ \ l}\Gamma_{\underset{\frown}{jl}}^{\ \ h} + \Gamma_{i\underset{\smile}{h}}^{\ \ l}\Gamma_{j\underset{\smile}{l}}^{\ \ h},\\ \Gamma_{i\underset{\smile}{j}} &= \Gamma_{i\underset{\smile}{j}}^{\ \ h}{}_{,h} - \frac{1}{2}(\Gamma_{i\underset{\smile}{h}}^{\ \ h}{}_{,j} - \Gamma_{j\underset{\smile}{h}}^{\ \ h}{}_{,i}) + \Gamma_{i\underset{\smile}{j}}^{\ \ h}\Gamma_{lh}^{\ \ h} - \Gamma_{i\underset{\smile}{h}}^{\ \ l}\Gamma_{l\underset{\smile}{j}}^{\ \ h} - \Gamma_{j\underset{\smile}{h}}^{\ \ l}\Gamma_{\underset{\frown}{il}}^{\ \ h}.\end{aligned}\right\} \tag{23}$$

For the connection $\Delta_{ij}^{\ \ h}$ we have in accordance with (11) and (14)

$$\Gamma_{\underset{\frown}{ij}}^{\ \ h} = \left\{\begin{matrix} h \\ ij \end{matrix}\right\}, \qquad \Gamma_{i\underset{\smile}{j}}^{\ \ h} = \Delta_{i\underset{\smile}{j}}^{\ \ h}, \qquad \Gamma_{i\underset{\smile}{h}}^{\ \ h} = \Delta_i = 0 \tag{24}$$

Substituting in (23) and noting that[3]

$$\Gamma_{\underset{\frown}{ih}}^{\ \ h} = \left\{\begin{matrix} h \\ ih \end{matrix}\right\} = \frac{1}{2}\frac{\partial \log g}{\partial x^i} \tag{25}$$

where $g$ is the determinant of $g_{ij}$, one obtains

$$\Delta_{ij} = -R_{ij} + \Delta_{i\underset{\smile}{h}}^{\ \ l}\Delta_{j\underset{\smile}{l}}^{\ \ h}, \qquad \Delta_{i\underset{\smile}{j}} = \Delta_{i\underset{\smile}{j}}^{\ \ h}{}_{/h}, \tag{26}$$

where $R_{ij}$ is the Ricci tensor[4] for the tensor $g_{ij}$ and $\Delta_{ij}^{\ h}/_h$ is the covariant derivative[5] of $\Delta_{ij}^{\ h}$ for the Christoffel symbols $\left\{ {i \atop jk} \right\}$.

From (21) one has

$$\Delta_{hj}^{\ h}{}_{k} = \Delta_{hj}^{\ h}{}_{,k} - \Delta_{hk}^{h}{}_{,j} = \left\{ {h \atop hj} \right\}_{,k} - \left\{ {h \atop hk} \right\}_{,j} = 0, \tag{27}$$

in consequence of (25).

If the generalized Riemann space is one for which $\Delta_{ij} = 0$, it follows from (26) that

$$R_{ij} = \Delta_{ih}^{\ l}\Delta_{jl}^{\ h}, \qquad \Delta_{ij}^{\ h}{}_{/h} = g^{hk}\Delta_{ijk/h} = 0. \tag{28}$$

The first of these equations involves linear terms in the second derivatives of the quantities $g_{ij}$. When the second is written in the second form it is seen from (12) that this equation involves linear terms in the second derivatives of $g_{ij}$. There are other such equations in view of the fact that the right-hand numbers of the first of equations (28) must satisfy the identities[6]

$$R^{h}{}_{j/h} = \frac{1}{2}R_{ij},$$

where

$$R^{h}_{\ j} = g^{hi}R_{ij}, \qquad R = g^{ij}R_{ij},$$

and $R^{h}{}_{j/h}$ is the covariant derivative of $R^{h}_{\ j}$ for the symbols $\left\{ {i \atop jk} \right\}$.

1. *R. G.*, p. 19. This is a reference to the author's *Riemannian Geometry*, Princeton University Press.
2. *R. G.*, p. 17.
3. *R. G.*, p. 18.
4. *R. G.*, p. 21.
5. *R. G.*, p. 28.
6. *R. G.*, p. 82.

---

## *WEIERSTRASS TRANSFORMS OF POSITIVE FUNCTIONS*

BY D. V. WIDDER

HARVARD UNIVERSITY

Communicated by Hassler Whitney, February 17, 1951

By the Weierstrass transform[1] we mean

$$f(x) = c\int_{-\infty}^{\infty} e^{-(x-y)^2/4}\varphi(y)dy, \qquad c = (4\pi)^{-1/2}. \tag{1}$$

This transform is of particular interest in the general theory of convolution transforms

$$f(x) = \int_{-\infty}^{\infty} G(x - y)\varphi(y)dy$$

studied recently by Hirschman and Widder.[2] The kernels admitted were

$$G(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{e^{xs}}{\Psi(s)}\,ds,$$

where the functions $\Psi(s)$ formed a subclass ($\gamma = 0$, below) of the *entire functions of type* II of Pólya and Schur,[3]

$$\Psi(s) = Cs^n e^{-\gamma s^2 + \delta s} \prod_{k=1}^{\infty}\left(1 - \frac{s}{a_k}\right)e^{s/a_k} \qquad (2)$$

$$(C \lesseqgtr 0,\ n \geqq 0,\ \gamma \geqq 0,\ a_k \text{ real}).$$

By the work of Schoenberg[3] it is clear that our kernels $G(x)$ form a subset of his *totally positive* functions. We have known for some time that our methods could be applied to the whole class, but that the results would be cumbersome in the absence of a workable real inversion operator for (1).

In the present note we restrict attention to the case when all factors of (2) disappear except $e^{-\gamma s^2}$, the one that was excluded from the earlier theory. Then $G(x)$, for a suitable choice of $\gamma$, becomes $e^{-x^2/4}$, the kernel of (1).

Eddington[4] showed formally that $e^{-D^2}$ is in some sense an inversion operator for the Weierstrass transform. However, his interpretation leads to divergent series for many of the functions (1). We now adopt a different definition which will turn out to be effective for the whole class of convergent transforms.

Let us show first why $e^{-D^2}$ may be expected to invert (1). From the Laplace transform

$$e^{x^2} = c\int_{-\infty}^{\infty} e^{-xy}e^{-y^2/4}dy$$

we have, after replacing $x$ by $D$ and $e^{-yD}\varphi(x)$ by $\varphi(x - y)$, that

$$e^{D^2}\varphi(x) = c\int_{-\infty}^{\infty} \varphi(x - y)e^{-y^2/4}dy = f(x),$$

from which

$$\varphi(x) = e^{-D^2}f(x).$$

Since $e^{-x^2}$ is not the Laplace transform of any function, we must look elsewhere for an integral interpretation of $e^{-D^2}$. We choose the following familiar formula

$$e^{-x^2} = c\int_{-\infty}^{\infty} e^{-y^2/4}\cos xy\,dy.$$

Using it as a guide, we adopt the following definition.

DEFINITION. *When* $0 < t < 1$.

$$e^{-tD^2}f(x) = ct^{-1/2}\int_{-\infty}^{\infty} e^{-y^2/4t}\{\cos yD\}f(x)dy,$$

$$\{\cos yD\}f(x) = \sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}y^{2k}f^{(2k)}(x).$$

*When* $t = 1$.

$$e^{-D^2}f(x) = \lim_{t\to 1} e^{-tD^2}f(x).$$

It is not our present purpose to exploit this new definition. We wish rather to use it for the statement of a result of the type of Bernstein's theorem[5] to the effect that all completely monotonic functions in an interval $(c, \infty)$ have a representation

$$f(x) = \int_0^{\infty} e^{-xy}d\alpha(y) \qquad \alpha(y) \uparrow .$$

Our result is the expected one; namely, that the class of functions (1), with $\varphi(y) \geqq 0$, is essentially characterized by the condition $e^{-tD^2}f(x) \geqq 0$. As usual in this type of Theorem, the class of functions considered must be slightly enlarged by changing (1) to the Stieltjes integral of an increasing function. The precise result follows.

THEOREM. *For a function* $f(x)$ *to be represented in the form*

$$f(x) = c\int_{-\infty}^{\infty} e^{-(x-y)^2/4}d\alpha(y),$$

*where the integral converges for all* $x$ *and* $\alpha(y)$ *is non-decreasing, it is necessary and sufficient that*

1. $f(x)$ *is entire*
2. $f(x + iy) = O(e^{-y^2/4})$ *uniformly in* $-R \leqq x \leqq R$ *for every* $R > 0$
3. $e^{-tD^2}f(x) \geqq 0 \qquad 0 < t < 1, -\infty < x < \infty$.

This Theorem will be proved elsewhere. It may be illustrated by the following pairs.

| $f(x)$ | $\alpha(y)$ |
|---|---|
| $x^2 + 2$ | $y^3/3$ |
| $e^x$ | $e^{y-1}$ |
| $e^{x^2}$ | $5^{-1/2}\int_0^y e^{u^2/5}du$ |
| $e^{-x^2/4}$ | $\begin{cases}0(-\infty, 0)\\ 1(0, \infty).\end{cases}$ |

[1] Hille, E., "Functional Analysis and Semi-groups," *A.M.S. Colloquium Publications*, **31,** 371 (1948). Pollard, H., "Integral Transforms," *Duke Math. J.*, **13,** 317 (1946).

[2] Hirschman, I. I., and Widder, D. V., "The Inversion of a General Class of Convolution Transforms," *Trans. A.M.S.* **66,** 135 (1949).

[3] Schoenberg, I. J., "On Totally Positive Functions, Laplace Integrals and Entire Functions of the Laguerre-Pólya-Schur Type," PROC. NATL. ACAD. SCI., **33,** 11 (1947).

[4] Eddington, A., "On a Formula for Correcting Statistics for the Effects of a Known Probable Error of Observation," *Monthly Notices Royal Astr. Soc.*, **73,** 359 (1914).

[5] Bernstein, S., "Sur les fonctions absolument monotones," *Acta Matematica*, **51,** 1–66 (1928).

# THE INTERCONVERSION OF AUDIBLE AND VISIBLE PATTERNS AS A BASIS FOR RESEARCH IN THE PERCEPTION OF SPEECH*

BY FRANKLIN S. COOPER, ALVIN M. LIBERMAN† AND JOHN M. BORST

HASKINS LABORATORIES, NEW YORK

Read before the Academy, October 10, 1950

In investigating the acoustic aspects of speech it has long been the practice to convert these extremely complex sounds into a visible display, and so to enlist vision as an aid in dealing with a problem which lies largely in the area of auditory perception. Of the various displays which have been used, perhaps the most effective is provided by the sound spectrograph, which has come to be recognized as a valuable research tool for the study of the acoustic correlates of perceived speech.[1] By examining numerous spectrograms of the same sounds, spoken by many persons and in a variety of contexts, an investigator can arrive at a description of the acoustic features common to all of the samples, and in this way make progress toward defining the so-called invariants of speech, that is, the essential information-bearing sound elements on which the listener's identifications critically depend. The investigator can also take account of the variations among spectrograms, and by correlating these with the observed variations in pronunciation, he can begin to sort out the several acoustic features in relation to the several aspects of the perception.

There are, however, many questions about the relation between acoustic stimulus and auditory perception which cannot be answered merely by an inspection of spectrograms, no matter how numerous and varied these may be. For any given unit characteristic of the auditory perception, such as the simple identification of a phoneme, the spectrogram will very often exhibit several features which are distinctive to the eye, and the information which can be obtained from the spectrogram is, accordingly, ambiguous. Even when only one feature or pattern is strikingly evident, one cannot be certain about its auditory significance, unless he assumes that those aspects of the spectrogram which appear most prominently on visual examination are, in fact, of greatest importance to the ear. That assumption, as we shall try to point out later in this paper, is itself extremely interesting, but it has not been directly tested, nor, indeed, has it always been made fully explicit.

To validate conclusions drawn from visual examination of spectrograms, or, more generally, to determine the stimulus correlates of perceived speech, it will often be necessary to make controlled modifications in the spectrogram, and then to evaluate the effects of those modifications on the sound as heard. For these purposes, we have constructed an instrument, called

a pattern playback, which reconverts spectrograms into sound, either in their original form or after modification.

The basic operating principle[2] is quite simple (Fig. 1). The playback scans a spectrogram from left to right along the time axis, using a line of light modulated by a tone wheel at some fifty harmonically-related frequencies which match approximately the frequency scale of the spectrogram. Those portions of the modulated light which are selected by the spectrogram—by transmission through a film transparency or by diffuse reflection from a painted design—are collected by an optical system and led to a phototube. Thus the photocurrent, amplified and fed into a loudspeaker, produces sounds which have, at every instant, approximately the frequency components shown on the spectrogram.

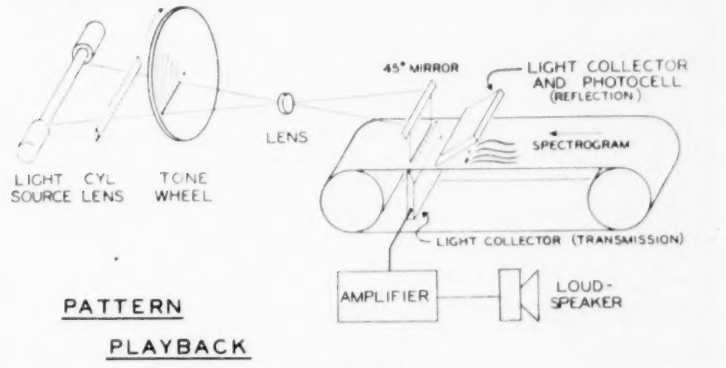



FIGURE 1

Operating principle of the pattern playback

For convenience in research the playback is designed to operate from either of two kinds of spectrograms (Fig. 2). In the one case, the spectrogram is a film transparency, and the sound is produced by the light which is transmitted through the relatively transparent portions of the film. These transmission spectrograms, so called, are photographic copies (on film) of an original produced from recorded sound by a spectrograph designed specifically for this purpose. The transmission spectrograms are most useful if one wishes to recreate the original sound as accurately as possible, or to make minor changes (especially deletions) in some detail of the spectrogram. In the other case, the playback operates from spectrograms which are drawn with white paint on a transparent plastic base, and the playback uses only the light which is reflected from the painted portions. The drawing is done by brush or pen, and the spectrograms can be prepared or modified in a variety of ways. These spectrograms are most

appropriate, then, if one wishes to make drastic changes in the sound, or, in the extreme case, to employ entirely synthetic patterns.

In general, the playback appears to be a most useful tool in research involving the experimental manipulation of speech sounds. By comparison with more conventional instruments for modifying the speech stream, the playback method is extremely flexible and convenient, and has the particular advantage that it allows considerable freedom in dealing with the dynamic or constantly varying aspects of speech. This was, indeed, the

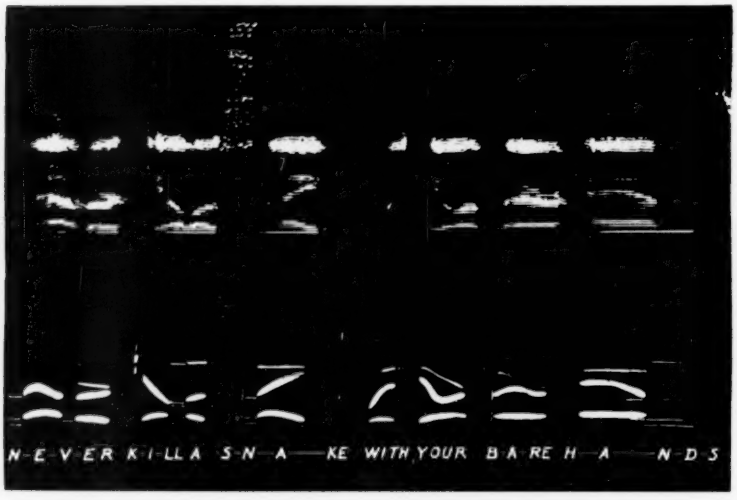



FIGURE 2

(*a*) Transmission spectrogram copied photographically from an original spectrogram without modification.

(*b*) Reflection spectrogram drawn by hand as a simplified version of the original spectrogram.

(*c*) Text of the sentence.

specific function for which it was designed. It is of course obvious that such a playback will be useful as a research tool only to the extent that it is able to produce intelligible speech which may then be degraded or dissected in various ways.

We have measured the intelligibility of the playback speech produced from transmission spectrograms, using for this purpose twenty standard test sentences, and have found it to be approximately 95%. The intelligibility of reflection spectrograms will obviously depend on how they are drawn. In drawing these spectrograms we have attempted, in a preliminary way, to determine the extent to which they could be simplified without

serious loss of intelligibility. The procedure was, first, to copy from an actual spectrogram the features which were most prominent visually, and then, largely on a trial-and-error basis, to make such further changes as were required for reasonable intelligibility. For the degree of schematization shown in figure 2 (*b*), the median intelligibility is about 85%, as determined with twenty test sentences. The experimental procedures and the nature of the simplifications will be reported in detail elsewhere.[3] For present purposes it is important merely to note that the simplified spectrograms are, in general, a reduction of the originals to their most obvious visual patterns. Although the first rough paintings were modified in many details in order to produce these highly simplified spectrograms, the modifications have not destroyed the over-all visual resemblance to the originals. To the extent that these similarities remain, and also to the extent that the simplified spectrograms are intelligible, these results provide a partial validation of the assumption, referred to earlier, that the spectrogram displays most prominently to the eye those acoustic features which are of greatest importance in auditory perception. If this were not the case, the spectrograph would not be so useful a tool in describing the sounds of speech, and, more significantly for our purposes, the playback would have no special advantage as a means of manipulating speech.

That the playback does have special advantages is indicated by our experience with it, and this fact seems, moreover, to have theoretical implications which deserve examination. It does not appear that the advantage is solely one of stopping time, that is, of converting a transitory sound into a stationary visible display which can be modified and then reconverted into sound for aural evaluation. This is an obvious advantage and an important one in experimenting with speech, but it is neither unique nor quite sufficient. For example, sounds can be represented to the eye by means of an oscillograph, and the oscillogram can be reconverted into sound by a device somewhat like a phonograph, yet the oscillographic representation is virtually useless as a basis for experimenting with the sounds of speech. The critical requirement, and the one which is not adequately met by the oscillogram, is that the display must provide for the eye information which is organized into patterns corresponding to the acoustic patterns on which auditory identifications depend; that is, the conversion must be from patterned acoustic information to patterned visual information. When this is so, the significant aspects of the acoustic pattern become comprehensible to the eye, and the display will have conceptual and also experimental utility in manipulating the sounds of speech. We believe that a reasonable approximation to the required conversion is represented by the spectrograph-playback combination, which interconverts the $x$ and $y$ coordinates of visual space with time and frequency in the acoustic domain (preserving intensity as a parameter in both cases), and that this

accounts for the special utility of these instruments as practical research tools.

On the theoretical side it would seem that the existence of this particular intersensory conversion, with its special advantages, may have interesting implications for the perceptual processes operating in vision and audition. Perhaps the most general implication is that there is an important similarity between visual and auditory perception, sufficient to permit an interconversion at the stimulus level such that patterns in the one sensory modality may, after conversion, be perceived as patterns in the other modality. The term pattern, as used here, implies an organization of stimuli which has the property of retaining its perceptual integrity in spite of gross and diverse changes in the absolute values of the several stimulus components. In vision, for example, a triangle will continue to be perceived as a triangle despite wide changes in size, position, etc. Similarly, in an auditory case, the perception of speech is to a certain extent independent of any fixed and isolated stimulus values, inasmuch as the sounds of speech can be considerably stretched or compressed in time, frequency and intensity without serious damage to intelligibility. These patterns of information, on which visual and auditory perception so largely depend, may then be defined operationally in terms of the particular alterations in stimulus which do and do not impair perceptual identification. By this definition, and according to our assumption about the general nature of the similarity between visual and auditory perception, it should be possible so to convert from vision to audition that varying the visual stimulus will affect the visual and the auditory identifications about equally, i.e., that stimulus changes which do or do not destroy the pattern in the one modality will, correspondingly, destroy it or not in the other. More simply, and only somewhat less precisely, stimuli which look alike can be made to sound alike, and stimuli which look different should then sound different.

It is most unlikely that visual and auditory patterns can be interconverted in complete detail and in all respects, and hence one must expect that the best of audio-visual transforms will be rather less than perfect. For all its inadequacies, however, such a transform will be of practical interest for its bearing on the problems of sensory prosthesis,[4] and for its application, as in the case of the playback, to investigations of the perception of a wide variety of sound patterns, including speech as a special and obviously important case.[5]

It appears, from a theoretical point of view, that the fact of intersensory pattern conversion implies functional similarities between vision and audition at a presumably high level of perception. Moreover, the specific nature of the similarities will be indicated by the precise way in which the best possible conversion is to be made. The transform concept also implies some degree of intersensory generality for certain perceptual laws, since

one may suppose that patterns will be interconvertible between vision and audition only to the extent that common principles determine the way in which visual and auditory stimuli are organized in perception. The further development and extensive testing of the transform concept, according to the criteria implicit in our assumptions, may therefore serve to reveal certain principles which are so basic as to have a special importance for theories about perceptual mechanisms in general.

We have taken a first and very tentative step toward testing the transform concept with materials other than speech sounds. In studying speech, we have been dealing with patterns which were quite obviously more

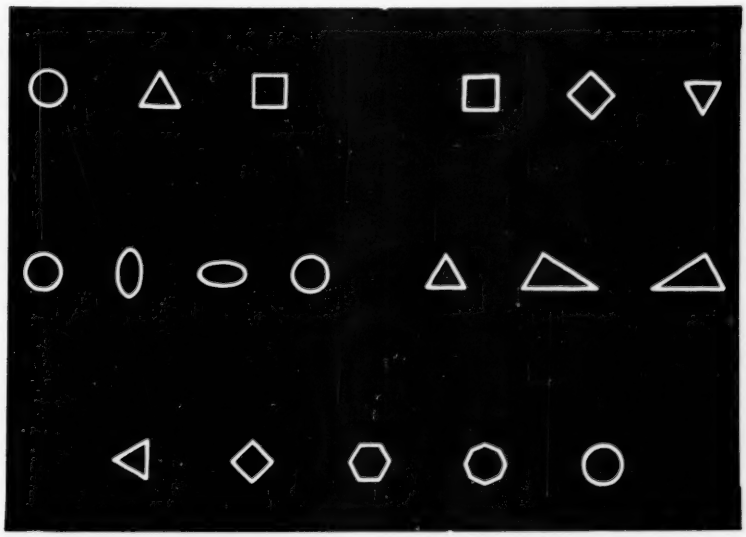


FIGURE 3

Three sequences of geometric figures used in testing the audio-visual transform.

familiar to the ear than to the eye. What will be heard when the playback is used to transform patterns which are familiar to the eye but not to the ear? The tests that we have so far made employed common geometrical forms such as circles, ellipses, triangles and squares (Fig. 3). As represented on the playback tape, each figure was varied in size and position, drawn in outline and in solid form, variously distorted in shape, and rotated about axes perpendicular to the tape. When these figures are converted by the playback, a listener hears sound which he rather readily puts into categories corresponding to the grouping of the figures in visual perception. All of the triangles, for example, sound somewhat alike, regard-

less of variations in size and position on the tape, and these sounds are quite distinctively different from the sounds corresponding to squares or circles. The one apparent exception is that rotation of many of the figures causes very large changes in the sound, and the listener tends not to put rotated forms of a rectangle, for example, into the same category. This may represent a limitation of the transform, though it must be remembered that the identifications of visual figures sometimes change when the figures are rotated, a familiar example being that of square and diamond.

These studies were intended to be exploratory and the results are most incomplete; there is not, of course, any special significance in the choice of geometric figures as test patterns. The experiments are described here because they illustrate how the playback transform can be tested according to our criterion of pattern interconvertibility, and because the results indicate in a general way the extent to which the playback transform, with its present imperfections, meets that test.

In summary, we have found it useful to assume, as a working hypothesis, that the perceptual processes in vision and audition exhibit important similarities in operating upon patterned information. A consequence of this hypothesis, put more explicitly and in operational terms, is that information which is perceived as a pattern when displayed to the eye can be so converted into sound for presentation to the ear that the pattern characteristics will be preserved, that is, that the sounds will be perceived as patterned information in audition. Conversely, patterned acoustic information can, by the reverse transformation, be displayed as visual patterns. The sound spectrograph and pattern playback are appropriate even though imperfect instrumental means of effecting such an interconversion, which in this case is that the $x$ and $y$ dimensions of the visual array are changed to the time and frequency dimensions of sound, with intensity remaining a parameter in both cases. Stated symbolically, the audio-visual transform then becomes $F_v(x, y, I) \rightleftarrows F_a(f, t, I)$.

This conceptualization of a particular relationship between vision and audition, quite aside from its theoretical interest, provides a basis for understanding the demonstrated utility of the sound spectrograph in displaying the sounds of speech for visual study, and of the pattern playback as a tool for the manipulation and synthesis of speech.

* The research reported here was made possible by funds granted by the Carnegie Corporation of New York. The paper, as read before the Academy on October 10, 1950, employed recordings to illustrate various points in the discussion. Some revision of the text has therefore been necessary.

† Also, University of Connecticut.

[1] Potter, R. K., and Steinberg, J. C., *J. Acous. Soc. Am.*, **22**, 803–823 (1950). Joos, Martin, *Acoustic Phonetics* (Language Monograph No. 23), Linguistic Society of America, 1948. Potter, R. K., Kopp, G. A., and Green, Harriet C., *Visible Speech*, D. van Nostrand, 1947.

[2] Potter, R. K., U. S. Patent No. 2,432,123 (1947). Cooper, F. S., Liberman, A. M., and Borst, J. M., *J. Acous. Soc. Am.*, **21**, 461 (1949). Vilbig, F., *Ibid.*, **22**, 754–761 (1950).

[3] Cooper, F. S., Borst, J. M., and Liberman, A. M., "Pattern Playback and Sound Spectrograph: Tools for Research in the Perception of Speech and Other Complex Sounds" (in preparation).

[4] Zahl, P. A., Ed., *Blindness: Modern Approaches to the Unseen Environment* Princeton University Press, 1950. Potter, R. K., Kopp, G. A., and Green, Harriet, C., *Visible Speech*, D. van Nostrand, 1947. Cooper, F. S., *Physics Today*, **3**, 6–14 (1950).

[5] Delattre, Pierre, *J. Acous. Soc. Am.*, **22**, 678 (1950). Cooper, F. S., Liberman, A. M., and Borst, J. M., *Ibid.*, **22**, 678 (1950). Cooper, F. S., *Ibid.*, **22**, 761–762 (1950).

# INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consecutively, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of Journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph., Boston*, **22**, 731–815 (1911); or, Wheeler, W. M., *Konigsberg, Schr. physik. Ges.*, **25**, 1–142 (1914).

MANUSCRIPTS may be sent to the Managing Editor, Edwin Bidwell Wilson, Harvard School of Public Health, 695 Huntington Ave., Boston 15, Mass., only by members of the ACADEMY or by chairmen of the divisions of the NATIONAL RESEARCH COUNCIL who will assume responsibility for the propriety and scientific standards of the paper. Articles are limited to six printed pages; *excess space is billed to the author at $14.00 per page.*

PROOF will ordinarily be sent,—with the understanding that charges for an author's corrections shall be billed to him. Authors are therefore requested to make final revisions on the typewritten manuscripts.

REPRINTS should be ordered at the time of submission of manuscript. They will be furnished to authors at cost, approximately as follows:

SCHEDULE OF RATES FOR REPRINTS, CARRIAGE CHARGES EXTRA

| No. Copies | 50 | 100 | 150 | 200 | 250 |
|---|---|---|---|---|---|
| 4 pp. | 4.13 | 4.72 | 5.30 | 5.90 | 6.50 |
| 8 pp. | 8.26 | 9.44 | 10.63 | 11.79 | 12.97 |
| Covers | 3.15 | 4.03 | 4.92 | 5.51 | 6.09 |

## CONTENTS


Page

CHEMISTRY.—ATOMIC COORDINATES AND STRUCTURE FACTORS FOR TWO HELICAL CONFIGURATIONS OF POLYPEPTIDE CHAINS . . . . . *By Linus Pauling and Robert B. Corey* 235

CHEMISTRY.—THE STRUCTURE OF SYNTHETIC POLYPEPTIDES . . . . . *By Linus Pauling and Robert B. Corey* 241

CHEMISTRY.—THE PLEATED SHEET, A NEW LAYER CONFIGURATION OF POLYPEPTIDE CHAINS . . . . . *By Linus Pauling and Robert B. Corey* 251

CHEMISTRY.—THE STRUCTURE OF FEATHER RACHIS KERATIN . . . . . *By Linus Pauling and Robert B. Corey* 256

CHEMISTRY.—THE STRUCTURE OF HAIR, MUSCLE, AND RELATED PROTEINS . . *By Linus Pauling and Robert B. Corey* 261

CHEMISTRY.—THE STRUCTURE OF FIBROUS PROTEINS OF THE COLLAGEN-GELATIN GROUP . . . . . *By Linus Pauling and Robert B. Corey* 272

CHEMISTRY.—THE POLYPEPTIDE-CHAIN CONFIGURATION IN HEMOGLOBIN AND OTHER GLOBULAR PROTEINS . . *By Linus Pauling and Robert B. Corey* 282

BIOCHEMISTRY.—THE INTERACTION OF MUCOPROTEIN WITH SOLUBLE COLLAGEN; AN ELECTRON MICROSCOPE STUDY . . . . . *By John H. Highberger, Jerome Gross and Francis O. Schmitt* 286

CHEMISTRY.—AN HYPOTHESIS ON PROTEIN SYNTHESIS POTENTIATED BY CITRIC ACID . . . . . *By Sidney W. Fox* 291

GENETICS.—AN ADENINE REQUIREMENT IN A STRAIN OF DROSOPHILA . . . . *By Taylor Hinton, John Ellis, and D. T. Noyes* 293

GENETICS.—ABSENCE OF CYTOSINE IN BACTERIOPHAGE $T_2$ . *By Alfred Marshak* 299

MATHEMATICS.—ON THE GENERATING FUNCTIONS OF TOTALLY POSITIVE SEQUENCES . . . . . *By Michael Aissen, Albert Edrei, I. J. Schoenberg, and Anne Whitney* 303

MATHEMATICS.—COHOMOLOGY THEORY OF ABELIAN GROUPS AND HOMOTOPY THEORY III . . . . . *By Samuel Eilenberg and Saunders MacLane* 307

MATHEMATICS.—GENERALIZED RIEMANN SPACES . . . *By Luther P. Eisenhart* 311

MATHEMATICS.—WEIERSTRASS TRANSFORMS OF POSITIVE FUNCTIONS . . . . *By D. V. Widder* 315

PSYCHOLOGY.—THE INTERCONVERSION OF AUDIBLE AND VISIBLE PATTERNS AS A BASIS FOR RESEARCH IN THE PERCEPTION OF SPEECH . . . . . *By Franklin S. Cooper, Alvin M. Liberman and John M. Borst* 318